
NATIONAL CENTER FOR EDUCATION STATISTICS

Conference Proceedings

FROM DATA TO INFORMATION
New Directions for the National
Center for Education Statistics

Gary Hoachlander
MPR Associates, Inc.

Jeanne E. Griffith
John H. Ralph
National Center for Education Statistics

U.S. Department of Education

Richard W. Riley
Secretary

Office of Educational Research and Improvement

Sharon P. Robinson
Assistant Secretary

National Center for Education Statistics

Pascal D. Forgione, Jr.
Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Office of Educational Research and Improvement
U.S. Department of Education
555 New Jersey Avenue NW
Washington, DC 20208-5574

August 1996

Suggested Citation

U.S. Department of Education. National Center for Education Statistics. *From Data to Information: New Directions for the National Center for Education Statistics*, NCES 96-901, by Gary Hoachlander, Jeanne E. Griffith, and John H. Ralph. Edith McArthur, project officer. Washington, D.C.: 1996.

Contact:

Project Officer
Edith McArthur
202-219-1442
FAX: 202-219-1575

Commissioner's Statement

In the fall of 1995, the National Center for Education Statistics (NCES) held a conference to stimulate dialogue about future developments in the fields of education, statistical methodology, and technology, as well as to explore the implications of such developments for the nation's education statistics program. This "Futures Conference" was unique for NCES because it attempted to combine considerations in all of these fields in order to stimulate the cross-fertilization and generation of ideas that might not emerge when discussing the topics separately. At this conference, the authors presented commissioned papers on targeted issues that were expected to be important over the next few years, and the discussants provided their comments.

From several perspectives, I believe the conference was highly successful. First, staff from NCES actively participated in all of the deliberations. As a result, they became personally engaged in the *process* of considering alternative futures for their agency. Since the "corporate culture" of this agency is to solicit and build on staff creativity, their participation and interest in this conference was vital. Second, both the formal and informal discussions generated many new ideas. The conference, as such, accomplished far more than the collection of commissioned papers alone could have because of the active interplay of ideas. Finally, many stakeholders in NCES's future saw this conference as a clear signal of the agency's commitment to continued improvement of the usefulness and quality of our surveys and data products. The stakeholders' positive response to the meeting was further reinforced by their expressions of interest in continuing to help in important ways. The success of the conference lies not in the sum of the individual presentations, rather in an overall perspective that provides guidance toward the future.

This publication will serve as a concrete reference to ensure that the stimulating ideas exchanged at the Futures Conference are not forgotten. While the quality of the discussion at the meeting was exceptional, one cannot expect to absorb everything said during a two-day conference. Thus, it is important to have a record that the participants can refer to this year, next year, or five years from now. Moreover, this publication will provide a way to share those ideas with others who could not participate in the conference. For instance, NCES has many customers and other stakeholders who have expressed keen interest in the conference proceedings and whose advice and considerations are welcome as a means to sustain the dialogue about NCES's future.

It is clear that if NCES wants to continue as a key player in providing information for education policy and decision making to the American public, policymakers, education researchers, and educators nationwide, it must continually reevaluate its program and products. In the future, we expect that NCES will receive requests for more of the kinds of products and services that it already provides. Also, we expect demands for new perspectives—on covering new topical areas, implementing new technologies, and adopting new methodologies. Already, major recent changes in the field of education are shaping our future program—for example, widespread innovations to

achieve education reform, efforts to adopt both curriculum and performance standards, and examination of education in the United States within an international context. Not only are methodological advances creating opportunities to produce statistics in ways that may be more efficient and effective, but also technological developments are changing the world in which we create data and disseminate our products more rapidly than ever before. The Futures Conference and this publication provide a new vision for NCES—a vision that acknowledges the constraints on the resources of governmental agencies at the end of the 20th century, as well as clearly emphasizes the opportunities that can be achieved with innovative methodologies and technologies and through close attention to the priorities for statistical knowledge in the field of education.

This contribution to envisioning NCES's future is occurring at a pivotal time of transition. The Futures Project was conceptualized under the leadership of the first Commissioner of Education Statistics, Emerson J. Elliott. It is our deep hope that this publication will serve as a tool for a new Commissioner to guide the agency in the upcoming years, as a source of ideas for planning and thinking, and as a foundation for long-term change in the organization.

Jeanne E. Griffith
Acting Commissioner

Acknowledgments

From Data to Information: New Directions for the National Center for Education Statistics began more than two years ago as a proposal from Jeanne Griffith. With the strong support and encouragement of Emerson Elliott, then Commissioner of NCES, Jeanne and staff in the Data Development Division guided and nurtured the project to its successful completion in the spring of 1996. John Ralph played a key role in shaping the focus and direction of the project, with important early contributions from Ron Hall, Ed Mooney, Paul Planchon, and Joe Conaty. Edith McArthur and Dawn Nelson also helped shepherd the project along.

Eighteen authors wrote papers that provided the project's substantive foundation and then presented their work to a national conference held in November 1995. Discussants from both inside and outside NCES also participated in this conference and contributed their expertise. *From Data to Information* assembles the thought, creativity, and analysis of these authors and discussants and summarizes some important considerations for policy, research, and practice. Without these individuals, the project would not have been possible, and they deserve special appreciation for the care and enthusiasm they brought to the undertaking. A complete list of authors and discussants is included in each of the project's publications.

At MPR Associates, Phil Kaufman helped mold the policy and methodological directions of the papers and conference discussions. Steve Klein and Patricia Holmes conducted a survey of leading educators and researchers in order to gather their ideas about the essential topics to be addressed by the project. Barbara Kridl was responsible for overseeing the production of the final product and editing it, Andrea Livingston for designing the volume and editing the draft, and Karyn Madden for editing the volume throughout its various iterations. Leslie Retallick and Denise Bradby created the cover, and Don Eike and Connie Yin produced the final product. Special appreciation also goes to Fena Neustaedter who provided ongoing administrative support, and to Laura Horn who offered many useful comments on the synthesis of the papers and conference proceedings.

To all of these individuals, thank you. Your work will figure prominently in discussions at NCES about both its current and future efforts to better inform the nation about education.

Gary Hoachlander
Project Director
June 1996

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

	Page
Commissioner's Statement	iii
Acknowledgments	v
1 From Data to Information: New Directions for the National Center for Education Statistics	
From Data to Information: New Directions for the National Center for Education Statistics	
Gary Hoachlander	1-1
Introductory Comments	
Emerson Elliott	1-28
2 Tracking Education Reform: Implications for Collecting National Data Through 2010	
Tracking Education Reform: What Type of National Data Should Be Collected Through 2010?	
John F. Jennings and Diane Stark	2-1
Where Are We Going? Policy Implications for Data Collection Through 2010	
Christopher T. Cross and Amy Rukea Stempel	2-12
Discussant Comments	2-19
3 Curriculum, Pedagogy, and Professional Development	
Enhancing Opportunity to Learn Measures in NCES Data	
Dominic J. Brewer and Cathleen Stasz	3-1
Teacher Education, Training, and Staff Development: Implications for National Surveys	
David R. Mandel	3-29
Discussant Comments	3-43
4 Trends in Statistical and Analytic Methodology: Implications for National Surveys	
“So What?” The Implications of New Analytic Methods for Designing NCES Surveys	
Robert F. Boruch and George Terhanian	4-1
Discussant Comments	4-116

Contents (continued)

	Page
5 New Data Collection Methodologies, Part II: Experimental Design	
Incorporating Experimental Designs Into New NCES Data Collection Methodologies	
Charles E. Metcalf	5-1
Discussant Comments	5-19
6 Postsecondary Education	
Tracking the Costs and Benefits of Postsecondary Education: Implications for National Surveys	
Michael S. McPherson and Morton O. Schapiro	6-1
Special Issues in Postsecondary Education and Lifelong Learning	
David W. Breneman and Frederick J. Galloway	6-13
Discussant Comments	6-29
7 New Data Collection Methodologies, Part I: Observational Strategies	
Large-Scale Video Surveys for the Study of Classroom Processes	
James W. Stigler	7-1
Discussant Comments	7-30
8 Education for Work: Curriculum, Performance, and Labor Market Outcomes	
Education and Work: Curriculum, Performance, and Job-Related Outcomes	
Peter Cappelli	8-1
Discussant Comments	8-35
9 Using Administrative Records and New Developments in Technology	
Administrative Record Opportunities in Education Survey Research	
Fritz Scheuren	9-1
New Developments in Technology: Implications for Collecting, Storing, Retrieving, and Disseminating National Data for Education	
Glynn D. Ligon	9-32
Discussant Comments	9-66
Appendix A: About the Contributors	A-1
Appendix B: Future NCES Data Collection Conference Agenda	B-1

1

From Data to Information: New Directions for the National Center for Education Statistics

THIS PAGE INTENTIONALLY LEFT BLANK

From Data to Information: New Directions for the National Center for Education Statistics

Gary Hoachlander

INTRODUCTION

In 1995, approximately 65 million Americans participated in elementary, secondary, or postsecondary education in the United States (U.S. Department of Education 1995b). Young people spent from one-fourth to one-half of their waking hours in school and school-related activities, while Americans of all ages continued to pursue some form of active learning that added to their repertoire of knowledge and skills. To serve these students, the nation's schools, colleges, and universities directly employed some 11.2 million people, even more than in the health industry. In addition, these direct services supported a substantial number of additional jobs in companies serving education through the production of everything from buses and computers to textbooks and software. All told, the nation spent more than \$500 billion¹ on formal education, or approximately 7.4 percent of Gross Domestic Product (U.S. Department of Commerce 1995). Probably no other single activity has occupied so prominent a place in family, community, and working life.

The primary purpose of the National Center for Education Statistics (NCES) is to describe this education enterprise and inform the nation about it. Congress charges NCES with collecting and reporting “statistics and information showing the condition and progress of education in the United States and other nations in order to promote and accelerate the improvement of American education.”² In doing so, the Center conducts a range of ongoing national surveys examining early childhood education, elementary and secondary education, postsecondary education, adult literacy, and the nation's libraries. Further, in cooperation with many other countries, it supports international surveys that aid in comparing educational progress and processes across nations. NCES carries out numerous analyses of these data and annually prepares more than 100 reports targeted toward policymakers, educators, researchers, and the American people.

Doing this job well would be necessary no matter what the focus, but when the subject assumes the magnitude and importance of education, this responsibility takes on special significance. Therefore, NCES must ensure that it continues to describe education fully and accurately and that it performs this function efficiently and thoroughly—in other words, that it remains well informed about important education issues, key advances in methods, and new developments in the technology of collecting, managing, analyzing, and reporting large amounts of information.

To this end, NCES undertook an in-depth examination of how best to direct its responsibilities for collecting and reporting information on education over the next decade. Three principal questions guided this effort:

- 1) What are the major issues and trends in education that NCES should aim to address through the first decade of the next century?
- 2) What are the most important advances in methods for collecting and analyzing information that should guide how NCES surveys are designed and used?
- 3) What opportunities do technological advances in data management and communications present for improving data collection and analysis and for disseminating findings and information effectively?

To help answer these questions, NCES conducted four activities: 1) a survey of leading educators and researchers, asking them to answer one or more of the three questions listed above; 2) commissioned papers addressing key topics suggested by the survey results; 3) a conference where the authors of the commissioned papers presented their work, with subsequent discussion by NCES staff and external reviewers; and 4) a published volume of the commissioned papers and discussants' comments. This paper summarizes and synthesizes the results of this work and consists of four major sections. This first introductory section provides a synopsis of the major themes and conclusions emerging from the papers and the conference. The second section describes the current foundation of NCES, delineating its core functions, operating principles, and program of work. The third summarizes some new directions that NCES could pursue to provide information for policy, research, and practice in American education, and also addresses some important methodological and technological opportunities. The paper ends with a brief conclusion.

Two dominant themes emerged from this collaborative effort. First, NCES must place greater emphasis on transforming raw data into information useful to policymakers, educators, researchers, and the general public than it does today. Accomplishing this goal will require that the relationships between NCES and data providers and between NCES and data users change significantly. During the next 5 to 10 years, the distinctions among these three parties—NCES, data providers, and data users—will become increasingly blurred, and their communications will probably become much more interactive, continuous, and two-way, with all three parties actively and simultaneously engaged in survey design, data collection, analysis, interpretation, presentation, and dissemination. Although technology will help pave the way for this transition, considerable conceptual thinking will also be required to take full advantage of the technological opportunities.

Second, in order to be more responsive to the demands for information about education, NCES will need to broaden its conception of what constitutes “data” and strategies for their collection. Traditionally, NCES has concentrated on designing and conducting surveys asking “closed-ended” questions that lend themselves to rapid, well-defined quantification. Although such surveys are likely to remain the hallmark of NCES's data collection activities for some time, the agency will need to pay more attention to how to supplement these data with various forms of “prequantified” material and observations. Technological developments will permit inexpensive collection of increasing amounts of textual, visual, and auditory data as integrated supplements to

surveys. This capacity should make it easier for researchers to ask questions and explore subjects that they did not foresee when designing the survey, thus enriching analytic power and reducing the expense of designing and conducting new surveys to examine unanticipated concerns.

In addition to these two major themes, this effort led to five important conclusions about future directions for NCES. First, NCES should strive to produce information that addresses more immediate and specific policy concerns. While the agency's role in monitoring and describing major long-term trends in education must not be compromised, this role will assume even greater importance if the agency can also contribute in a timely way to more focused policy debates. The widespread emphasis on education reform during the past 10 years has spawned a large number of different strategies for improving education. As a result, policymakers at all levels—national, state, and local—want to know more about what has and has not been accomplished.

Second, surveys yielding better information that bears directly on the practices of teaching and learning would significantly enhance the contribution of NCES to both policy debate and research. Since current surveys produce scant data on the specific content of curriculum, the nature and frequency of discrete classroom activities, the practices of teachers, or the kinds of tasks students perform in order to learn, the classroom remains largely a “black box” that defies clear understanding and precise strategies for improvement. Without a clearer understanding of what constitutes effective classroom practices, it will be difficult to do more than simply describe what kinds of education reforms have been implemented. Whether they have, in fact, improved teaching and increased learning will remain unknown.

Third, survey designers should consider more carefully strategies that will permit integrated analysis of the interrelationships among education inputs, processes, and outcomes. Although existing surveys do an excellent job of providing nationally representative descriptive data on many important aspects of education, they do not, however, lend themselves very well to reliable causal analyses that might increase knowledge about what works and why. The descriptive power of national data must be preserved, but there are promising new designs emerging, which, if selectively incorporated into national surveys, might generate more robust conclusions about the relative effectiveness of various educational practices.

Fourth, NCES should make better use of data already collected and maintained by others. Doing so will help NCES simultaneously accomplish three aims: 1) expand the amount and type of data it collects; 2) adopt a wider range of data collection and analytic methods; and 3) function within the tight resource constraints that are certain to affect almost all federal agencies. Previously, NCES has pursued such a strategy with some success—for example, through the Common Core of Data (CCD) for elementary and secondary education and the Integrated Postsecondary Education Data System (IPEDS); however, high standards for data quality, especially comparability and reliability, have frequently forced the agency to collect new data that were already available in a somewhat different form or from a different time period. Without doubt, NCES must maintain its data quality standards, but increasing cooperation with states and localities, combined with rapidly improving data management technology and communications, should create opportunities for the Center to do a better job of streamlining and coordinating data collection.

Fifth, NCES will need to place increasingly greater emphasis on dissemination. Data and information are only as valuable as the breadth, quality, and timeliness of the uses made of them. Electronic storage media (data tapes and compact disks) and printed publications will surely remain the cornerstone of the agency's strategy for distributing data, tabulations, and the results of analysis. However, NCES should pay more attention to clearinghouse and brokerage functions, as well as effective use of electronic networks.

These themes and conclusions do not represent radical departures from the major path NCES has been pursuing in recent years. Indeed, as the following section illustrates, they are well suited to building on the foundation of core functions, operating principles, and programs of work that support the current agency. Nevertheless, serious attention to these ideas will almost certainly produce important differences in what the agency now does and how it does it.

BUILDING THE FUTURE ON THE CURRENT FOUNDATION

In 1986, the Panel to Evaluate the National Center for Education Statistics, a group created under the auspices of the National Academy of Sciences (NAS), reported on the results of its 2-year assessment of the mission and effectiveness of NCES (Levine, ed. 1986). The Panel was created to address the widespread perception that the existing agency had not yet developed “the image and the reality of a competent and objective major statistical organization serving the wide need for statistics about education in the United States” (Levine, ed. 1986, p. 13). To address such problems as quality of data, timeliness, conceptual obsolescence, and insufficient funding and staff, the Panel made many important recommendations, including the following:

- Clearly establish and define the Center's role in ensuring the availability of data needed to describe the condition of education in the United States;
- Improve the compilation of education program, staff, and financial data from the states, including developing closer collaboration with the states to ensure that the Center's program of work meets both NCES and state requirements for usefulness, relevance, quality, and reliability;
- Strengthen the Center's methodological and technical capacity through more systematic use of outside expertise in the Advisory Council of Education Statistics, as well as ad hoc advisory groups;
- Develop, publish, disseminate, and implement standards to guide all phases of the Center's work, including establishing an office of statistical standards headed by a chief statistician;
- In collaboration with the states, assess and improve the quality, consistency, and reliability of data obtained from state and local agencies, from institutions of higher education, and from other sources; and
- Institute a publications policy that clearly distinguishes between different types of reports—for example, statistical summaries and digests, analytic reports, descriptive reports, and reports on methodology—and develop a schedule of fixed release dates for selected key education statistics.

Ten years later, with the direction provided by the Panel, strong leadership at NCES, and support from Congress and the larger education community, NCES is much stronger and has become a widely respected statistical agency. The agency has significantly strengthened its core functions; operates under well-defined guiding principles and high standards for data collection, analysis, and reporting; and has established a clear program of work for reporting on the major aspects of education in the United States and other nations. Whereas 10 years ago, the future of NCES depended on rectifying fundamental weaknesses, today the agency's future can build on a strong foundation.

The Core Functions of NCES

The National Academy of Science's Committee on National Statistics defines the principal purpose of a federal statistical agency as “the compilation and analysis of data and the dissemination of information for statistical purposes” (Martin and Straf, eds. 1992). NCES adheres to this primary purpose by organizing its work around three core functions:

- 1) Survey Design and Data Collection
- 2) Information Production—data analysis, translation, and interpretation
- 3) Dissemination

Since national surveys are the primary means for NCES to collect data on education, during the past 10 years, the Center has devoted much effort to improving survey design and data collection. For instance, following the recommendations of the 1986 NAS Evaluation Panel, it has developed and implemented various strategies to improve data quality, to detect and reduce error, and to expedite data collection. The Center has also significantly improved the sophistication and efficiency of its sampling methods, increased its use of computer-assisted telephone interviewing, and systematically assessed the quality of data generated in its national surveys.

Moreover, NCES has strengthened and substantially expanded its capacity to analyze data. Rigorous statistical standards now govern all aspects of its analytic function, from simple tabulations to the most sophisticated multivariate analyses.³ The Center routinely applies procedures for quality control to all of its surveys, which include analyzing data quality, eliminating unacceptable error, and producing methodological and descriptive summary reports before releasing survey data for public use.

Finally, NCES has greatly expanded and improved its dissemination function. Toward this end, the agency has developed and implemented publication standards that now guide the production of NCES reports and the release of public use data files. A central new feature of the agency's dissemination function has been developing strict policies for protecting the privacy of participants in NCES surveys. The Center not only applies safeguards to the data released to the public but also requires that its analytic contractors follow strict requirements for limiting access to prerelease data files and for maintaining the confidentiality of survey respondents. Failure to adhere to these requirements carries stiff fines, as well as the possibility of imprisonment.

Operating Principles

In carrying out these core functions, NCES adheres to three operating principles:⁴

- 1) Produce information that is policy relevant, while maintaining strict impartiality, institutional independence, and neutrality with respect to programmatic effectiveness;
- 2) Maintain credibility with users of its data, analysis, and publications; and
- 3) Maintain trust among those who provide data, including individuals, institutions, and public and private agencies.

NCES's program of work must be guided by the issues and requirements of public policy and federal programs, while scrupulously avoiding specific policy recommendations or identification with particular policy agendas or ideological perspectives. This principle is perhaps easiest to achieve when the Center performs its responsibilities for providing data to others for analysis or when it produces tabulations and descriptive reports. When NCES engages in analysis or interpretation, however, it must exercise greater care to remain policy neutral while still contributing relevant information to policy debates.

Attention to this principle has important implications for charting future directions for NCES. The suggestions that the Center address more immediate, specific policy concerns and that it develop survey designs more strongly suited for evaluation of what works could lead it beyond the boundaries of policy relevance into policy statements and evaluation. This, in turn, could jeopardize its position of impartiality. Deriving greater policy benefit from data and information produced by NCES, therefore, must proceed with great care.

Attention to this first principle also contributes to realizing the second, credibility with users of NCES data and information. However, credibility depends on more than policy relevance and impartiality. It also derives from confidence in the rigor of survey design, the quality of the data, the strength of analysis, and the accessibility and usability of its products, publications, and services. Here again, as NCES considers making greater use of data collected and maintained by others, it will need to guard against undermining its credibility with users who now depend on the Center's increasing emphasis on methodological rigor and data quality.

Finally, the success of NCES as an information agency rests on the trust it engenders among those who supply it with data. Protecting the privacy of survey participants is a key aspect of maintaining this trust, and integrating new types of data into national surveys will pose challenges for assurances of confidentiality. Use of video and audio data—for example, taping teachers in the classroom—will require close scrutiny of this issue. Confidentiality, however, is not the only condition for securing trust among data providers. Suppliers of data also need to be confident that the information being requested is truly needed, that it will be tabulated and analyzed accurately, and that providers will be given opportunities to correct errors or clarify ambiguities. Pressures for greater timeliness or more direct electronic access to decentralized, raw data files may undermine the confidence of data providers in the absence of explicit attention to new strategies and safeguards.

Program of Work

NCES has organized its current program of work around seven major topics:⁵

- 1) Elementary and Secondary Education
- 2) Postsecondary Education
- 3) Educational Assessment
- 4) National Longitudinal Studies
- 5) International Comparative Studies
- 6) Vocational Education
- 7) Libraries

Information on each of these topics is produced from a variety of surveys and studies, several of which supply data to more than one topical area. Some of the surveys, such as the CCD (on elementary and secondary schools and school districts), are designed as a census of the universe of respondents, which then serves as a sampling frame for more in-depth cross-sectional or longitudinal surveys on smaller samples of the population. The Schools and Staffing Survey (SASS), for example, collects detailed information on teachers and administrators in a sample of schools drawn from the CCD.⁶ In other instances, a large comprehensive survey provides the basis for a more intensive study of a subset of respondents. In this vein, the National Postsecondary Student Aid Study (NPSAS)—a nationwide survey of students enrolled in postsecondary institutions—provides the basis for more targeted longitudinal studies of students who are starting postsecondary education, the Beginning Postsecondary Students (BPS) Longitudinal Study, and of students who have completed a baccalaureate degree or higher, the Baccalaureate and Beyond (B&B) Longitudinal Study.

Central to the NCES program of work are various surveys and studies designed to assess the knowledge, skills, and performance of American students. For instance, the National Assessment of Educational Progress (NAEP), which conducts assessments of reading, mathematics, writing, science, history, and geography for samples of students enrolled in elementary and secondary education,⁷ is probably the best known of these efforts. In addition to NAEP, NCES also provides other data on student performance through transcript studies (at both the secondary and postsecondary levels); through the National Adult Literacy Survey, which examines adults' ability to use prose, documents, and mathematics in a variety of commonplace daily activities; and through international assessments that provide comparative information about student performance in the United States relative to that of other countries.

Finally, the Center conducts several long-term longitudinal studies designed to track students' paths through school and into subsequent stages of working and family life. These have included such studies as the 1980 High School and Beyond (HS&B Study), the National Education Longitudinal Study of 1988 (NELS:88), and the Early Childhood Longitudinal Study (ECLS), which is still in the planning and testing stage and is expected to begin with a kindergarten class in 1999.

These surveys now contribute to approximately 100 publications that NCES produces each year, including descriptive reports, analysis reports, methodological reports, issue briefs, and a variety of other documents. Three of these documents—the *Digest of Education Statistics*, *Projections of Education Statistics*, and *The Condition of Education*—annually provide a broad national overview of education at all levels in the United States.

In summary, during the past 10 years, NCES has been engaged in a process of steady development and improvement. In 1996, NCES is a viable and credible statistical agency, applying high standards to the provision of information on the condition of education in the United States and the nation's progress toward improving mastery of knowledge and skills among all its citizens. With this strong foundation, the agency is now well positioned to pursue some new directions that will enhance its ability to produce important information for policy, research, and practice in American education.

NEW DIRECTIONS IN INFORMATION FOR POLICY, RESEARCH, AND PRACTICE

In considering how NCES can best chart a course over the next decade that will capitalize on the foundation of work already in place, it is useful to consider its contribution to three domains of education: policy, research, and practice. These domains are by no means mutually exclusive; in fact, they overlap and interact in important ways. There are, however, information needs that are either unique or more dominant in each, and it is therefore instructive to consider the following questions individually:

- How can NCES best contribute information to discussions of education policy at the national, state, and local levels?
- How can NCES contribute information that will support significant research on education effectiveness and improvement?
- How can NCES contribute information that supports practice—i.e., the “front-line” activities that develop knowledge and skill in the nation's students?

A fourth question constantly weaves through these first three: what advances in methodology and technology can assist NCES in providing useful information to each of these domains? This section addresses each of these four questions in turn.

Information for Policy

The agenda of NCES is, in the first instance, greatly influenced by public policy issues and the requirements of federal, state, and local programs affecting education. Information contributing to policy debates can assume at least three major forms:

- 1) *System indicators* that describe the functioning of the education enterprise, broadly and over the long term;

- 2) *Implementation indicators* that describe the breadth and depth of the execution of policies and practices; and
- 3) *Effectiveness indicators* that describe the results achieved by students and educational institutions and programs.

Although system indicators have been a long-standing focus of NCES, there are potentially important new developments for the agency to consider. NCES surveys have included information on the implementation of some generic policies and practices, but specific federal and state policy initiatives have not been examined. Surveys have also included measures of student outcomes—the NAEP is the best known example; however, these measures typically cannot be directly linked to particular policies or educational practices to permit rigorous assessments of effectiveness. What are some possible new directions for NCES to consider with respect to each of these three types of indicators?

System Indicators

Data that portray the major aspects of the American education enterprise, both cross-sectionally and over time, form the core of the mission and functions of NCES. Reporting basic descriptive information on students, faculty and other staff, institutions and governing districts, and education finances must continue to be the primary focus of NCES and should not be compromised by new initiatives. The authors contributing to this examination of new directions for NCES are unanimous on this point: the primary purpose of NCES is and should remain representatively describing and documenting the condition of education in America and other nations.

This basic description of the education system can, of course, be improved, and several of the papers included in this volume offered suggestions.⁸ Among the kinds of system information the authors would like to see developed are the following:

- Detail on curriculum content, including rigor and substance;
- Detail on the nature and frequency of particular teaching practices, especially those which research indicates are effective;
- Attention to the nature and frequency of student behavior that reflects engagement in learning;
- Resource allocation at the institutional and classroom level;
- Measures of teacher quality and the ways in which teachers apply their knowledge and skills in the classroom;
- More contextual information on postsecondary institutions, especially their objectives in awarding student financial aid and improved coverage of proprietary institutions;
- More attention to the interaction between education and work; and
- More attention to governance issues, particularly new organizational and oversight arrangements.

In many respects, these recommendations represent requests for “finer grain” in the descriptive data presently collected by NCES. In some instances, this aim can be accomplished by asking for more detailed information; in other instances, collecting and reporting existing data at lower levels of aggregation (the classroom, for example, rather than the school or school district) will be necessary.

Implementation Indicators

The widespread attention on education reform during the past 10 to 12 years has spawned a number of new policy initiatives at the national, state, and local levels. Congress periodically revises such mainstay education legislation as the Elementary and Secondary Education Act, the Higher Education Act, or the Individuals with Disabilities Education Act. Additionally, it undertakes new education policy initiatives such as GOALS 2000 or the School-to-Work Opportunities Act. States have also initiated many new policies to strengthen elementary, secondary, and postsecondary education. These have included changes in the requirements for high school graduation, new teacher certification regulations, modifications to postsecondary admissions standards, and new policies on college tuition and student financial aid.

Traditionally, NCES has not monitored the implementation of specific federal or state legislation. At the national level, Congress has typically provided for independent assessments or evaluations of education legislation, such as the National Assessments of Vocational Education, which have been conducted approximately every 5 years. While these national assessments make extensive use of NCES data, they also conduct independent surveys that focus more particularly on key features of the legislation being examined.

Several of the authors involved in this project have urged NCES to monitor some of the key policies and practices that have emerged from federal, state, and local legislation during the 1980s and 1990s.⁹ It should be emphasized that they are not recommending that NCES assume responsibility for evaluating particular legislation, because they believe this function should continue to rest elsewhere. Rather, they are urging NCES to examine policies and practices that became more generic as they have been adopted and implemented through various federal, state, and local initiatives and are, therefore, no longer associated with any single piece of legislation. Some specific examples include the following:

- Curriculum content standards and measures of student or institutional performance;
- Length of the school day or year;
- Requirements that students complete particular courses (for example, in math, science, or foreign language) or accumulate a minimum number of credits for graduation;
- Participation in a variety of “work-based” learning opportunities, including apprenticeship, cooperative education, tech-prep programs, or school-based enterprise;
- Operation of charter schools;
- Prevalence and nature of home schooling;
- Availability and use of school choice;

- Participation in reform networks, such as the Coalition of Essential Schools, Accelerated Schools, or *High Schools That Work*;
- Changes in affirmative action policies;
- Changes in postsecondary admission requirements;
- Prevalence of state takeovers of local school districts or other forms of state intervention in financially troubled localities; and
- Changes in state policies affecting postsecondary tuition or student financial aid.

More attention to such issues by NCES would help ensure that its data are policy relevant, while still leaving responsibility for policy evaluation to independent studies and other agencies in the Department of Education.

Effectiveness Indicators

In addition to information on how to implement policies and practices, policymakers would also like better information on their effectiveness. Even though it is useful to know how widespread the adoption of a particular strategy for improving education has been, it is even more useful to know how well it has worked, and why or why not. This, of course, is a primary aim of most policy evaluation, as well as many research projects.

Much of the credibility of NCES rests on its clear separation from policy evaluation and research on education impacts and outcomes. Although NCES contributes essential data and information to these efforts, it remains well removed from the conduct of any of these activities. This separation of functions contributes to the neutrality and objectivity that NCES must maintain as the nation's primary statistical agency for education. The impartial character of NCES must be preserved. Consequently, any initiative to make NCES surveys more conducive to assessments of policy effectiveness must proceed with great care.

Why consider such a course at all? First, there is potentially a substantial payoff from better integrating the nationally representative features of NCES surveys with the more rigorous but also more narrowly circumscribed designs of policy impact studies. In the current environment, policy analysts often face a frustrating choice: asking the right question with weak methodology and data that were not collected specifically for that purpose, or asking a much less important question with sound methodology and specially tailored information.¹⁰ Clearly, answering important questions with sound methods and precise information is more likely to improve education and the policies that support it. Combining the representative power of national surveys with the methodological rigor of experimental design would help realize this objective.

Second, there may be significant cost savings from integrating some impact evaluations with national surveys. Both kinds of efforts are quite costly. It is not unusual for a national survey to cost in excess of \$10 million, and the more rigorous policy evaluations adopting experimental design frequently cost as much or more. Both efforts often collect similar kinds of data at approximately the same points in time, sometimes even from the same respondents. Eliminating this duplication

would not only reduce costs but also alleviate some of the burden on respondents participating in national surveys and evaluations.

Cost savings aside, the primary benefit of integrating methodologically rigorous effectiveness assessments with nationally representative surveys lies in increasing the usefulness of these two activities beyond the results obtained when they are conducted independently. National survey data would more directly and authoritatively address questions about policy effects; impact studies would be conducted in a nationally representative context that would increase the likelihood that study results could be generalized.

To achieve this result, NCES should carefully consider piloting the inclusion of an experimental study in one of its national surveys. Any of the longitudinal surveys now under way are potential candidates, including SASS, ECLS, or the longitudinal spin-offs of NPSAS.

What should be the focus of experimental studies imbedded or linked to national surveys? Clearly, the choice must be considered carefully, with ample input from interested policymakers, researchers, and educators. Given the mission of NCES, focusing on a particular type of educational practice would probably be more appropriate than on an assessment of a specific legislative program. One possibility, for example, would be to conduct a careful study of the consequences of homogeneous versus heterogeneous grouping of students by academic ability.¹¹

Information for Research

Researchers are heavy users of information produced by NCES, and while the boundaries between policy and research are fuzzy, the interests of the research community deserve some separate attention. In the papers produced for this project, three themes emerged as priorities for focusing NCES's contribution to research over the next decade:

- 1) Teaching and Learning—illuminating more clearly what actually happens in the classroom;
- 2) Education Production—clarifying the processes of transforming education resources into student, program, and institutional outcomes; and
- 3) Education Outside the Classroom—depicting what and how learning occurs beyond the walls of the traditional classroom in homes, workplaces, and the community at large.

Though not exhaustive, this list provides some important directions for NCES to consider. In the next section, each topic will be briefly discussed.

Teaching and Learning

Much of the business of education occurs in the nation's classrooms—elementary, secondary, and postsecondary—yet national surveys presently tell us relatively little about what actually takes place at the classroom level. Currently, good information is available about the different types of courses taught (at both the secondary and postsecondary levels), but there is little or no nationally

representative detail on the content of the curriculum or how it varies among classrooms, institutions, or states. Similarly, not much data are available on teaching practices, either the range of strategies adopted by faculty or the frequency of their use. Finally, most surveys do not offer much description of what students do to facilitate or impede learning in the classroom.¹²

While richer information on these three aspects—curriculum, pedagogy, and student behavior—would be useful in and of itself, the greatest benefit to research is likely to be achieved when information on all three is simultaneously available at the individual classroom level. That is, ideally researchers would want to examine how these three aspects of classroom activity interact and to understand how they relate to various types of education outcomes. In this way, more can be learned about what works and why in the daily business of education.

To realize this objective, one implication for future NCES surveys is clear: survey designs need to pay more attention to using the classroom as a unit of analysis. Additionally, the designs should strive to produce an *integrated* package of information on curriculum content, teaching practices, student behaviors, and student learning outcomes. It is not sufficient, for example, to simply expand transcript studies to include more information on course content; rather, expanded information on course content must be linked to other data on teaching practices, student behavior, and student achievement.

In addition to the question of what kinds of information on classroom activity can best advance future research, there is also the issue of how best to collect it. Traditionally, to obtain data on classroom activities, NCES has asked respondents questions through paper questionnaires or telephone interviews. Thus, to the extent that current surveys yield information on teaching practices or student behavior, they rely mainly on self-reports.

An alternative to collecting information through respondent self-report is direct observation by trained researchers. Until recently, direct observation has been a very expensive alternative, indeed prohibitively so for large-scale surveys involving thousands of respondents. However, recent technological and methodological advances are making direct observation, as well as the collection of source materials, more feasible.¹³ Video is one of the most promising strategies for linking direct observation to more traditional survey techniques, and NCES is using this technique for the first time in designing the Third International Math and Science Survey (TIMSS).¹⁴

Video, of course, is not an especially new technology. What is new, however, is its rapidly growing capacity to store large amounts of video information inexpensively in digital form that enables fast retrieval and analysis. Additionally, researchers are making steady progress in developing analytic techniques that simplify and accelerate transforming video information into coded data suited for analysis using quantitative methods. Video, therefore, can add significantly to the richness and analytic potential of a survey, since it reduces the need to anticipate all of the questions the survey must ask of respondents. As researchers observe video records, they can formulate completely new variables that may not have been considered in the design phase of the survey. In the past, such new formulations usually required asking respondents follow-up questions or designing a new survey, if such avenues were pursued at all.

Closely related to this kind of use of video technology is the increasing capacity to collect, store, and analyze large amounts of textual information. For example, if researchers need better information on the content of textbooks or other printed materials used in classrooms, it is now possible to optically scan samples of these classroom documents for subsequent coding and analysis. As with video images, electronically storing and retrieving large amounts of textual information is relatively inexpensive.

These advances in storing and analyzing large amounts of what is essentially “prequantified” data promise to integrate survey research with case study methods, and represent research strategies that until now have been pursued independently of one another, each with its own strengths and weaknesses. This integration has the potential to link the representative statistical power of survey design with the richness and variety of case study information, simultaneously obtaining the best of both worlds.

Expanding surveys to include systematic collection of prequantified visual, textual, and even auditory information could produce significant new contributions to research on teaching and learning. Consequently, NCES should carefully consider how best to capitalize on its initial experience with this strategy in TIMSS, with special attention to adapting the approach to other surveys such as SASS, ECLS, or the longitudinal spin-offs of NPSAS. Additionally, the use of video in national surveys might prove especially beneficial if it were initially combined with efforts to imbed experiments in national surveys. The combination of these two methods targeted on analyzing the effectiveness of particular teaching interventions, for example, could yield very useful and robust results.

Education Production

Better understanding the interactions among curriculum, pedagogy, and student behavior in the classroom is an important piece of a larger set of research questions—how dollars are allocated (to localities, institutions, and classrooms), transformed into various resources, organized into programmatic and teaching strategies, and used to produce increases in students' knowledge and skills. In short, NCES data could play a much more significant role in expanding knowledge about how to better use education resources to improve student performance, thereby improving the overall process of education production.¹⁵

Achieving this goal will require some changes in the way NCES currently collects data on the financing of elementary, secondary, and postsecondary education. At present, there are two main surveys collecting financial data, CCD at the elementary and secondary level and IPEDS at the postsecondary level.¹⁶ With respect to financial information, both of these surveys focus primarily on providing detail on revenues and expenditures, for local school districts in the case of CCD and for individual institutions in the case of IPEDS. Both surveys are designed to collect financial data primarily from an accounting perspective and are not now well suited for cost-benefit analysis of educational programs or cost-effectiveness analysis of particular teaching strategies. Neither provides information on the allocation of resources at the classroom level.

Providing data that better inform understanding of the production process of education would be aided by NCES expanding its present focus on finance to embrace a broader concentration on the

economics of education. This larger conception would aim to integrate data on finance with other data on education processes and practices, as well as student outcomes. Additionally, new kinds of economic data would be required. Rather than needing more detail on expenditures for such general functions as administration, instruction, maintenance, or capital outlay, researchers would want to obtain data on the costs of specific types of staff, different kinds of school improvement strategies, alternative teaching strategies, and so on. They would also want to know more about the costs of different kinds of course content, equipment, instructional products, and assessment. In short, rendering the process of education production more intelligible depends on moving beyond traditional concerns about the distribution and expenditure of dollars to a more careful examination of how to transform dollars into effective teaching and learning in the classroom.

Three strategies for improving NCES data on education finances would help accomplish this goal. First, what constitutes useful financial data needs to be reconsidered, with special attention to better information on unit costs and transforming dollars into education processes and practices. Second, data will be needed at the classroom level; information on districts or institutions is not likely to contribute much to this kind of research. Third, it must be possible to link these financial data to other data on teacher characteristics, classroom practices, student demographics and behavior, and learning outcomes. Without this kind of integrated data about how education occurs, understanding more precisely how to efficiently allocate resources for education will continue to elude researchers and policymakers.

Education Outside the Classroom

Although elementary, secondary, and postsecondary classrooms are the centers of formal education in America, it is widely understood that much learning also takes place outside the classroom in the home, the workplace, and the community at large. However, we know relatively little about what or how learning occurs in these settings, nor do we know much about how learning in these places interacts with learning in the classroom. Moreover, given that most Americans spend only 12 to 16 years in formal schooling but another 50 years or so learning in these informal environs, a thorough description of the condition of education in America would require closer attention to the learning that transpires beyond classroom walls.

NCES surveys have already paid some attention to nonschool settings. At the present time, probably the largest of such efforts is the Early Childhood Longitudinal Study (ECLS), which will begin by focusing on the preschool lives of a cohort of children who will be followed over their early years of development. Additionally, other longitudinal studies, such as HS&B and B&B, have collected data on respondents' experiences in the workplace. Information on labor market participation, however, has been limited primarily to data on types of labor market outcomes—for example, earnings, duration of employment, and types of occupation—rather than systematically examining how learning occurs in the workplace or the degree of congruence between learning goals in schools and education requirements on the job.

Comprehensively surveying learning that occurs outside the classroom is a tall order, and NCES should approach this task incrementally. One place to focus an expanded examination of informal learning is on the workplace and the strategies adults use to maintain and upgrade the knowledge and skills needed to remain productive, actively engaged workers.¹⁷ Such a focus is more

important than ever, given the changes that are occurring in today's work world. These changes include not only rapidly developing new technologies but also major shifts in the attachments and relationships between employers and employees. As the likelihood of lasting employment with a single employer becomes increasingly tenuous in the modern economy, individuals must assume ever greater responsibility for nurturing their own careers and continuing employability. How working adults will meet this responsibility in the future poses important new challenges for the nation's systems of education.

Increased attention to learning through and for work could begin with the following steps. First, it is important to learn more about the knowledge and skills needed for long-term success in the labor market. Are these requirements consistent with the academic and vocational goals of formal education, and how well does the formal education system produce the desired prerequisites? Second, NCES could pay closer attention to how learning occurs in the workplace; whether the process differs in important ways from learning in the classroom, and whether the two complement or reinforce one another. Third, NCES could enrich work-related data in its current longitudinal surveys, concentrating especially on better information about what people do on the job, what contributes to their successes and failures, and how they use or do not use school-based learning to perform and advance.

As part of its own mission, NCES could independently address all three of these issues. Alternatively, the agency may want to explore opportunities for collaborating with the Department of Labor and its surveys of employers and employees. Data collected by the Labor Department tend to provide greater detail on labor market participation, while being relatively deficient on education variables. Better coordination or integration of the two Departments' survey efforts could yield some important benefits.

Information for Practice

Most teachers and administrators are accustomed to viewing data as something to be reported to others. For example, they report daily attendance to central offices to document federal, state, and local funding systems. They submit grades for report cards to students and parents and for recording on student transcripts, which in turn are reported to postsecondary admissions offices. They administer standardized achievement tests for state assessments and college admissions. Rarely, however, do teachers and administrators use data directly themselves to improve their own programs and practices. One consequence of this outlook on data is that most practitioners do not make much use of the information provided by NCES. An important challenge for NCES, therefore, is significantly increasing the value and utility of its data for local teachers and administrators.

At least three strategies for providing better information for practice offer some important opportunities for NCES:

- 1) *Benchmarking*—helping local practitioners make comparisons against established norms;
- 2) *Networking*—linking practitioners with other practitioners and helping them discover more quickly who is doing what and where;

- 3) *Practitioner-Based Research and Self-Reflection*—engaging practitioners in systematic inquiry through NCES surveys and related research.

Benchmarking

“How well am I performing?” This is an appropriate question for any professional concerned with improving practice and increasing students' mastery of knowledge and skills. For most educators, however, it is a difficult question to answer in any way other than in an impressionistic or anecdotal fashion. Until recently, education has not had enough success with helping schools, programs, and faculty to monitor their accomplishments or to use the results to improve what they do.

Fortunately, as more and more states and localities develop new strategies for tracking performance and promoting school improvement, this situation is changing. “School report cards” are now produced annually in many states. Other states have developed systems of performance measures and standards, along with procedures for school improvement plans in districts that perform below state norms. Moreover, “keeping score” and using the results to assess the relative effectiveness of different kinds of school improvement strategies are core operating principles of several large consortiums, such as *High Schools That Work* under the auspices of the Southern Regional Education Board.

NCES could make an important contribution to the continued development of these practices by improving the utility of its survey results as benchmarks for states and localities interested in knowing how their performance measures up in relation to others. A local school or school district, for example, could find out how well its record on student attendance or high school completion compares with a national or state norm. It could then further refine the comparison by examining such measures in a subset of districts or schools that are similar with respect to size or student demographics. In addition to making comparisons at a particular point in time, a local school or school district might also monitor its relative performance over time. For example, is its success in reducing dropout rates proceeding at a faster or slower pace than in comparable districts or schools?

There is nothing to prevent localities, or even individual teachers, from using current NCES data to establish these kinds of benchmarks. However, they need to work rather hard to do it. Finding the right data is not always easy, nor is determining whether the NCES estimate is comparable with a local statistic. Tailoring an NCES estimate to yield a comparison of “likes with likes” requires a knowledge of NCES data sets, as well as analysis techniques, that most practitioners do not have. Thus, there are significant barriers to transforming NCES data into useful benchmarks at the local level.

However, there are at least three steps that NCES could take to make benchmarking easier for states and localities. First, in collaboration with potential state and local user-practitioners, NCES could systematically review its current dissemination activities with specific attention to how some aspects of the dissemination process could be modified to facilitate benchmarking. For example, there may be consensus on a relatively small set of indicators that NCES could publish annually in a succinct, accessible form with widespread local distribution. Such a publication might be similar

to *The Pocket Condition of Education*, which NCES now produces annually, but it could be designed with local benchmarking specifically in mind.

Second, NCES could explicitly consider local benchmarking when designing selected surveys, including customized reporting of results to survey participants. At present, survey participants receive little or no direct benefit from taking part in NCES surveys, and the burden of doing so is often not trivial. Providing participants with a summary of where they stand on selected variables relative to others in the survey could be a useful service.¹⁸ Such a summary could take the form of a traditional printed report. Alternatively, NCES might want to explore new electronic strategies that could actually distribute some limited analytic capacity along with the data (see discussion below on technological innovations).

Third, as NCES increases its capacity grows to provide information “on-line,” it should consider strategies for developing and distributing analytic packages that enable state and local benchmarking. In other words, instead of simply making data available, NCES would also provide a menu of data analysis programs or routines that would enable practitioners to generate their own statistics quickly and easily, without requiring a sophisticated knowledge of the underlying methodology. Such a strategy would build on NCES's current practice of providing users with “table generators,” increasing both the kinds of analysis that users could perform and the ease of using the analytic software.

Networking

Local teachers and administrators often want to know who else has experience with a particular school improvement strategy, type of curriculum, or teaching practice. Yet, systematically locating and communicating with other knowledgeable practitioners can be quite difficult; often it is not easy to find out who these individuals are or how to contact them. If NCES were to assume a greater role in monitoring the implementation of more specific education policies and practices (see earlier section on Implementation Indicators), it could also facilitate networking among practitioners. In addition to providing practitioners with information about the frequency with which a particular reform is being implemented and where it is being attempted, NCES could also match up interested parties and help them share information about their experiences.

This kind of knowledge brokering would represent a new function for NCES, one that may not be completely in keeping with traditional perspectives on the appropriate role of a statistical agency. Nevertheless, as NCES develops its presence on the Internet and the World Wide Web, this kind of service would be an obvious extension of its capacity to transform data into information valued by practitioners. Moreover, when providers of data also have a direct use for similar information from others, they are much more likely to respond to NCES's requests in an accurate and timely fashion. Thus, NCES's ability to monitor implementation for policy purposes could well be enhanced by its also using the information to provide an important service to teachers and administrators.

Practitioner-Based Research and Self-Reflection

Until recently, surprisingly little has been known about the specific elements of high-quality teaching (and by extension, high-quality teachers). This lack of knowledge has contributed to much misinformation and misunderstanding about what it takes to be a good teacher, as well as confusion in the public policy arena over the role of professional development in education reform. The status of national data on teachers reflects this state of affairs, with facts limited mainly to demographic characteristics and scant information available on the quality of practice or practitioners.

A very promising development, therefore, is the recent effort on the part of the teaching profession to begin a systematic, sustained examination of what constitutes good teaching—specifically what teachers should know and be able to do to help students master high levels of proficiency. Exemplified by the work of the National Board for Professional Teaching Standards (NBPTS), as well as other organizations and state-level initiatives, this effort is forging some consensus about appropriate standards for defining advanced high-quality teaching. This work has led to the establishment of a voluntary system of certification for early childhood, elementary, middle, and secondary school teachers, including differentiation among a range of academic disciplines (for example, math, science, history and the social sciences, English, and vocational education).

These developments create an important opportunity for NCES not only to improve the data it gathers on the nation's teachers but also to contribute more directly to strengthening teaching.¹⁹ This opportunity can be realized in two ways. First, as the work on teaching standards and certification continues to evolve, NCES should be able to define a larger array of indicators of teaching quality to include in national surveys. At a minimum, these indicators should focus on measuring teachers' command of the knowledge bases and teaching methods that are being identified as reflective of high-quality practice. Gathering such data could rely on traditional methods of written assessment or self-report. Alternatively, if NCES opts for further developing video observation techniques, these methods could significantly enrich information on the condition of teaching nationwide.²⁰ Furthermore, as more teachers choose to pursue national certification and as more certificates are awarded, national counts of teachers participating in and successfully completing the process will assume greater value as indicators of teacher quality.

Second, as NCES pursues this first strategy for improving data on the quality of practice, it could actively engage practitioners in this process and create opportunities for more interactive research and development. For example, if written examinations (in the style of NAEP) or video observation become part of NCES's strategy for monitoring and reporting to the nation on teacher quality, this process could be designed to simultaneously benefit individual teachers participating in the surveys. This might be accomplished in several ways. The results of written assessments could be returned to individual teachers. Groups of teachers could be assembled to review and constructively critique video segments. Further, if the data gathering process included collecting mini-portfolios submitted by teachers, these could be systematically evaluated, with examples of best practices culled from the data and disseminated to teachers and teacher education institutions. In short, the business of collecting national data could begin to play a more direct role in professional development and the strategic improvement of teaching and schools.

Further Considerations About Methodology and Technology

Increasing the contribution of NCES to policy, research, and practice depends in part on closer attention to a number of methodological and technological opportunities. Some of these, such as imbedding experimental designs in national surveys or collecting prequantified data through video and optical scanning, have already been discussed. There are some additional considerations, however, that deserve special mention, including: 1) developments in using administrative records, 2) promising techniques for obtaining hard-to-get information or producing more finely tuned estimates, and 3) effective use of the Internet and World Wide Web.

Administrative Records

Much of the information sought by national surveys already exists, at least in an approximate form, in records maintained for administrative purposes by schools, postsecondary institutions, district offices, state agencies, and other public and private offices. Transcripts, for example, provide detailed information on courses attempted and completed, grades, credits earned, and scores from standardized tests. Personnel records contain data on teaching assignments, salaries, demographics, qualifications, and experience. And budget and accounting offices maintain extensive records on revenues and expenditures. To the extent that surveys can access and use these administrative records, they often can obtain information that is more accurate than the responses provided by survey participants, often at significantly less cost.

Several NCES surveys already rely heavily on administrative records for information. Some good examples are CCD, IPEDS, and the NPSAS. There are, however, two types of problems that have limited the usefulness of administrative records. First, the contents of the records may not meet acceptable standards of accuracy, consistency, and comparability. Second, access to administrative records is often problematic, for a variety of reasons ranging from concerns about confidentiality to technical problems that may be as mundane as locating the right filing cabinet in the right office.

Technological advances in computing and electronic networking promise to reduce both of these problems considerably over the next decade, and NCES should be alert to opportunities to exploit new developments.²¹ First, electronic administrative records maintained in easy-to-use relational databases will increasingly become accepted practice among the nation's elementary, secondary, and postsecondary education systems,²² since they will be building administrative databases to satisfy their own needs and uses for data. Consequently, *collecting* data will less often be viewed as an externally imposed burden and cost. Whether *providing* these data to national surveys will be seen as burdensome, however, will depend critically on the ease with which data can be transmitted to those requesting information. To facilitate transmission, NCES will need to pay particular attention to assisting with the standardization of data elements and with the development of cheap scannable forms and other strategies for promoting electronic access and transfer.

Second, as local educators and administrators become more sophisticated users of data (rather than just providers), the business of designing surveys, collecting and analyzing data, and reporting results is likely to become much more interactive. The traditional model in which NCES assumes primary responsibility for all of these functions is likely to yield to much more decentralized, distributed models in which the respective roles of surveyor and respondent become less distinct and

more intertwined. For example, respondents who are also users of data may play a much greater role in defining survey questions and data elements. They may also develop specialized analyses (including analytic routines) that are shared with other respondent/users. NCES, in turn, may assume more responsibility for coordinating and brokering surveys, analyses, and reporting, rather than unilaterally directing and conducting all of these activities.

One possible implication of these trends is a reversal in the respective roles of independent surveys and administrative records in providing national data for education. To date, administrative records have mainly been adjuncts to large-scale surveys; they have supplemented data collected through written questionnaires or telephone interviews. In the not-too-distant future, administrative records may become the basic building blocks of national data systems, with smaller targeted questionnaires designed as supplements.

Methods for Producing Better Statistical Estimates

To provide good information for policy, research, and practice, NCES relies on a wide variety of survey design, data collection, and analytic methods—some relatively simple and widely known, others extraordinarily complex or reflecting recent advances in specialized fields. For purposes of this synthesis, a thorough discussion of survey and analytic methods is neither possible nor appropriate. The papers produced for this project, however, raised and discussed a number of methodological issues and developments. Four of these deserve special mention for careful consideration by NCES in charting its future course.²³

First, as mentioned previously, NCES should exploit opportunities to combine well-designed, targeted controlled experiments within the national surveys that have been the hallmark of its data collection activities. These experiments must be compatible with the mission and conduct of the larger survey effort, and a particular experiment should not be undertaken if it risks jeopardizing the nationally representative and descriptive power of the survey in which it is imbedded. However, if these criteria can be satisfied, imbedded experiments are promising examples of constructing a “whole exceeding the sum of its parts.” Such experiments could contribute significantly to knowledge about what works and why in the nation's classrooms.

Second, survey questions that elicit information on sensitive topics must always be considered with great care. National surveys about education are no exception, and it is important that NCES does not avoid issues simply because they are sensitive or controversial. Methodological developments can help reduce some of the concern surrounding this issue. For example, one promising strategy called “network-based estimating,” in which respondents are asked about the behavior of unidentified acquaintances in their social network, has been developed by quantitative anthropologists. There is growing evidence that this procedure produces indirect but reliable information on sensitive topics, without depending on the respondent to report directly on his or her own personal experience. NCES should explore the feasibility of using this or similar techniques in future surveys.

Third, and related to the second issue, there are promising new developments in methods for generating indirect estimates of statistics at subnational levels or for intervening periods of time between surveys. Traditionally, producing estimates for smaller units of analysis—states, for example, or institutions within states—has depended primarily on increasing sample size. Similarly, obtaining estimates more frequently—say, every 5 years rather than every 10—typically requires administering the survey more frequently. Both of these strategies are usually quite expensive. An alternative method being developed uses auxiliary data (from ongoing administrative records, for example) along with the survey data to produce indirect estimates for smaller units of analysis or intervening periods of time. Successfully adapting these techniques to some NCES surveys could yield more finely grained estimates at a modest cost.

Fourth, there are long-standing calls for better linking and integrating the databases produced by NCES surveys. As noted earlier, a better understanding of how learning occurs in the classroom will require simultaneous access to data on curriculum content, teaching practices, student behavior, and student outcomes. It has been rare to find a single database with rich information on all of these attributes for a sufficiently large sample, however. If NCES were able to significantly improve the connections between its surveys, it is likely that opportunities for better, more focused research would be enhanced. However, in order to achieve this long sought-after objective, NCES must do substantial work, both conceptually and methodologically, to determine precisely what is meant by “linking” and “integrating.”

Finally, NCES must continue to actively promote methodological developments and adaptations suited to its mission. Most researchers, whether engaged in a particular substantive pursuit or methodological advancement, are occupied primarily with their own interests and agendas; they are not paying much attention to the relevance of their work for NCES. Consequently, NCES needs to provide for the orderly acquisition and screening of methodological and technological applications to surveys and analysis. There are many strategies for doing this, including advisory groups, grants, conferences, commissioned papers, and so on. Whatever strategy is chosen, however, the basic objective must be an explicit and high-priority item on the agenda of NCES.

Internet and the World Wide Web

No discussion of future technological developments would be complete without some mention of the Internet and the World Wide Web.²⁴ However, the pace and variety with which these are evolving make any effort to forecast precisely their role in the work of NCES quite difficult, if not simply foolish. Perhaps, the most useful approach is to use the evolution of the information highway as a metaphor for changes in communication and interaction between NCES and the public it serves. In some respects, the Internet and the World Wide Web will facilitate and hasten these changes, but in others, they are simply reflective of larger forces at work in contemporary society.

Presently, NCES is in the second stage of a three-stage evolution in how many organizations typically interact with their clients. In Stage One, to accomplish its mission, NCES dominates the relationship between itself and those who either provide or use the data it gathers. Communication

tends to be mainly one-way and follows well-established paths. NCES designs the surveys, administers questionnaires or interviews, collects and cleans data, conducts its own analysis, and produces and disseminates reports. Other analysts of NCES data pursue their research independently; they do not feed back results to NCES, at least in any systematic fashion. Stage One loosely represents the pre-Internet, pre-Web world. It is history.

In Stage Two, which coincides with the advent and initial development of electronic networks, relations between NCES and providers and users of data become more interactive (though still predominantly one-way), and in some instances the distinction between data *provider* and *user* begins to blur. In this stage, surveys begin to make more use of electronically stored administrative records, and, consequently, questions or data elements may be tailored to a particular respondent. Selected providers and users may be authorized limited on-line access data to update or correct information. Although NCES continues to generate substantial analyses on its own, it also begins to pay closer attention to the analytic objectives of users. In addition to distributing data files, it also disseminates analysis files designed to facilitate specific types of research—the relationship between education and labor market participation, for example. Additionally, NCES may provide analysts with software to accelerate their analyses or to ensure that those who conduct external analyses of NCES data adopt appropriate statistical techniques. Reports are made available in electronic form, and specialized electronic user groups or technical review panels begin to form on the network. Currently, NCES is already well immersed in Stage Two.

In Stage Three, which will emerge more clearly and strongly with greater access to electronic networks and with deeper understanding about how to use them effectively, relationships between NCES and its data providers and users will become truly two-way and continuous. Any data user, who could also be a data provider (a state office, for example), might send NCES a small software program that initiates a customized database search, adds the results to NCES's data library, and returns a tailored report to the original requester. Conversely, NCES may be constantly developing small software programs that go out over the network and retrieve data needed to respond to specific inquiries from Congress, researchers, educators, or the public at large.²⁵

Surveys may assume the form of database development, with specifications designed interactively by users and providers coordinated by NCES. The scale of written questionnaires or telephone interviews will diminish considerably or be limited to highly focused inquiries. Much of this design process will occur on-line through electronic conferencing among NCES, data users, and data providers. Even though NCES will probably still produce many of its own reports, electronic versions of these documents will contain numerous electronic links to other data sets, technical references, and related reports. They may also contain interactive software that will permit users to perform “what if” analyses while perusing a report and to generate customized tables or graphics. Alternatively, users will generate their own electronic reports and analyses and create links to NCES documents residing in electronic networks. Precisely what is generated by NCES and what is generated by others may become less easy to distinguish.

Stage Three is not here yet, and it will probably look quite different from this admittedly inchoate prediction. However, this stage will probably arrive much sooner than expected. The more NCES can anticipate and help shape these developments, the more likely it will be able to use them effectively to report on the condition of education in the United States and other countries.

CONCLUSION

From data to information—transforming quantitative facts about education into knowledge useful to policymakers, researchers, practitioners, and the general public—this aim has always been central to the mission of the National Center for Education Statistics. In and of itself, this objective is not a new direction for the agency. However, what constitutes useful information and how it gets produced, distributed, and used are changing. To keep pace with these changes, indeed to stay out in front and help shape their development, NCES must chart some new directions.

Probably the most fundamental change that NCES will need to address is its emergence as a provider of information *services and systems*, rather than a primary collector and provider of data per se. In today's climate of growing demands for information, but limited resources to produce it, NCES will need to pay particular attention to assuming new roles as a facilitator, broker, translator, linkage, filter, and pathfinder in a complex web of providers and users of education data. To these new roles, the agency can bring a strong foundation of standards for high-quality data and analysis, as well as a firm understanding of the kinds of information that are most relevant to deliberating national policy for education.

As these new roles develop, NCES may find itself shedding or at least de-emphasizing old functions. Data collection that occurs independently of front-line administrative and teaching systems and their own information needs is likely to diminish significantly. This change, in combination with technological advances, may lead to data collection systems that are far more decentralized, interactive, and operating in “real time” than the systems that have traditionally supported national surveys. It is even possible that eventually NCES may find that it is no longer in the data collection business, as this function has traditionally been defined. Instead, it will be primarily a systems manager and analyst, a producer and broker of information for ongoing nationally oriented assessments, as well as thousands of state and local customized queries. Data collection and storage, however, may occur largely outside of the immediate domain of NCES.

“Reporting statistics and information showing the condition and progress of education in the United States and other nations in order to promote and accelerate the improvement of American education”—this charge is a lasting mission for the National Center for Education Statistics. Fulfilling it successfully will require careful attention to changing national priorities, a strong commitment to improving education research and practice, and an openness to recognizing and adopting important advances in methods and technology.

NOTES

1. U.S. Department of Education, National Center for Education Statistics, “Common Core of Data” and “Financial Statistics of Institutions of Higher Education,” surveys and unpublished data, FY 94–95.
2. Section 402(b) of the National Education Statistics Act of 1994 (20 U.S.C. 9001).
3. Indeed, there are now standards for all of NCES major activities—survey planning and testing, statistical processing, data provision and analysis, evaluation and documentation, and contract management and operations. See U.S. Department of Education (1992).
4. These follow the principles for a federal statistical agency developed in Martin and Straf, eds. (1992).
5. For a full description of the current program of work at NCES, see U.S. Department of Education (1995a).
6. SASS also surveys private school teachers in a sample of schools drawn from the Private School Universe file maintained by NCES.
7. At present, legislation requires that NAEP assess reading and mathematics every 2 years; science and writing at least every 4 years; and history, geography, and other subjects determined by the National Assessment Governing Board at least every 6 years.
8. In particular, see the papers by Brewer and Stasz, Mandel, McPherson and Schapiro, Breneman and Galloway, and Cappelli.
9. See especially the papers by Jennings and Stark and by Cross and Stempel.
10. See the papers by Metcalf and by Boruch and Terhanian.
11. See the discussion in the paper by Boruch and Terhanian. These authors also suggest that NCES consider adopting a “satellite” policy that would permit including controlled experimental studies in national surveys in a fashion similar to the way NASA allows adjuncts to space missions for astrophysicists and others.
12. See the papers by Brewer and Stasz and by Mandel.
13. For a summary of trends in technological capacity to store, retrieve, and analyze data, see the paper by Ligon.
14. See the paper by Stigler.
15. See the paper by McPherson and Schapiro.

16. While SASS provides additional data on salaries of administrative and instructional personnel at the elementary and secondary level, NPSAS provides additional postsecondary information on tuition and costs.

17. See the paper by Cappelli.

18. There are important confidentiality considerations that must be addressed if this kind of service were to be provided. However, with explicit attention to benchmarking at the outset of a survey, problems surrounding confidentiality could be reduced.

19. See the paper by Mandel.

20. Video already plays an important role in the certification process used by NBPTS, and NCES could build on the experience of the Board, as well as that of other researchers developing this technology.

21. See the papers by Ligon and by Scheuren.

22. Electronic recordkeeping is still far from universal, especially at the elementary and secondary levels; paper files are still the norm in many places.

23. These are developed in more detail in the papers by Boruch and Terhanian, Metcalf, and Scheuren.

24. For more information about specific opportunities for NCES to use electronic networks, see the papers by Boruch and Terhanian, Ligon, and Scheuren.

25. The continuing development of “object technology,” a technique for more rapidly constructing software programs out of many small object modules, should hasten the explosion of this sort of interactive dissemination and sharing of programs.

REFERENCES

- Levine, D. B. (Ed.). 1986. *Creating a Center for Education Statistics: A Time for Action*. Washington, DC: National Academy Press.
- Martin, M. E., and Straf M. S. (Eds.). 1992. *Principles and Practices for a Federal Statistical Agency*, p. 2. Washington, DC: National Academy Press.
- U.S. Department of Commerce. 1995. *Statistical Abstract*, table 699, p. 451. Washington, DC: Bureau of the Census.
- U.S. Department of Education. 1992. *NCES Statistical Standards*. Washington, DC: National Center for Education Statistics.
- U.S. Department of Education. 1995a. *Programs and Plans: 1995 Edition*. Washington, DC: National Center for Education Statistics.
- U.S. Department of Education. 1995b. *Digest of Education Statistics 1995*, table 3, p. 12. Washington, DC: National Center for Education Statistics.

Introductory Comments

Emerson Elliott

The authors of papers described in this volume have performed their services to the National Center for Education Statistics thoughtfully and with a serious purpose. As a collection, these examinations of possible future directions for education statistics are valuable—although not in laying out a plan of action for the Center, because the Center's leadership must do that. Rather, their value is in sketching a vision for a federal statistical information service and in showing that new technologies, statistical methodologies, and more powerful analytic procedures can achieve that vision even in a time when government is under pressure to achieve more with less.

Why is that so special? It was my frequent experience, during the years I was with NCES, to solicit advice about what the Center should be doing through commissioned papers or conferences. What data should it gather? What were the issues about which policymakers, educators, and the public needed information? And precisely what information from which sources (e.g., students, teachers, parents, schools, school boards) would answer those questions? Center staff have made aggressive efforts to keep abreast of technology in data gathering and analysis, and following criticisms from a National Academy of Sciences evaluation in 1986, the Center made numerous changes to assure that appropriate methodologies and rigorous quality controls were being employed in its statistical work. But those experiences, together, led me to conclude that the Center's now expanded budget and program activities had reached their limit in recent years. These new papers, though, while certainly assuming at least modest continuing growth in the Center's program, also indicate a potential for achieving much more statistical information for the investment.

My comments first address the sense of vision that flows through these pages and then turn to three implications of that vision—on technology and administrative records, research and statistics, and analysis.

VISION

The major theme expressed in the papers is that the Center should serve as an education information agency, not just a statistical office. This perspective is especially prominent in the papers prepared by Jack Jennings and Diane Stark and by Christopher Cross and Amy Stempel, although it is implicit in the others as well. These papers described an agency responsive to policymaker interests in education information about a variety of current topics, that would be gathered in different forms (some of them costly, such as longitudinal studies) and that would make the resulting information available in accessible forms. The Center would, among other things, be a source of

information about the progress of reform even if that information is descriptive and not “statistical” in either the random survey sense or in the systematic extracting of data from traditional administrative records.

For example, Jennings and Stark emphasize in their paper that the Center should provide information on standards-based reform, on charter schools, on state course-taking requirements, on “reform networks” (such as those of Levin and Slavin), on choice programs, and on private, for-profit, companies that provide educational services.

Cross and Stemple, similarly, according to the Center should provide data on student mobility, school safety, moral and character education, technology, home schooling, charter schools, magnet schools, vouchers, and site-based management.

Thus, Jennings and Cross and their colleagues cast the Center not as a “data collection agency” but as a federal office with a broader education information function. It reminds me of the decline in the American railroads following World War II. Railroad executives viewed themselves as being in the railroad business, and lost out as transportation for individuals came to be defined by interstate highways, widely affordable cars, new airports, and jet engines. They never recovered from their own restricted vision of the railroad business.

Data collection is a tool that provides information. But that is too narrow a conception as questions turn to what the data say. The challenging question of what the data *say* can only be answered through an analysis function, through decisions about specifically which data will tell us something important, and through attention to relationships in data that have been examined and explained in research. Moreover, the very meaning of “data collection” is changing as electronic data systems come into widespread use, creating the possibility of moving information from administrative records instantaneously from place to place. The new potential of electronic capability is much more information at little cost, including data from “small areas” (schools, even classrooms, rather than districts or states or the nation) that are difficult and expensive to reach with traditional statistical methods.

At the same time, realizing this new vision would impose a change on the role of NCES as a statistical agency. The Center would “own” less data, since the information would be distributed around the nation at sponsor sites, frequently school districts, states, or institutions of higher education. The Center could capture the data for summarizing and analytic purposes, but others could as well—and surely would. As Glynn Ligon put it, the line between data retrieval and data dissemination would be blurred. The Center's role in building consensus for terms and definitions and common means of access, already exercised, would grow, probably with support and encouragement from the nation's educational and governmental institutions. They, too, will want access to data in some predictable way and will find the resulting statistics of greater use if they have the same meaning from place to place and can be connected from elementary and secondary education to higher education.

This vision is very attractive because it would make the Center relevant, I think essential, in the 21st century. It would, however, require developing connections and consensus about what is useful that involve more partners more consistently than has been the practice up to now.

ON RESEARCH AND STATISTICS

This group of papers describes roles for the National Center for Education Statistics that would require activities well beyond the bounds of traditional statistical agencies. Let us start with some brief definitions. I think of “statistics” as data that are designed to respond to those instances when information is to be representative of a population—when you want to know how something is distributed among individuals in different situations; when you want to know what happens over time; and when you want to know if a relationship derived from theories, experiments, and case studies will hold up when you “go to scale.” “Research” also produces data, but is driven by a need to understand qualitative relationships, a need for in-depth information, or an intent to evaluate consequences of a specific intervention. There is no reason that policymakers should be required to make these distinctions—that should be done right here in the Office of Educational Research and Improvement. And it is certain in this collection of commissioned papers that the authors did not attempt to sort out what NCES should do and what other parts of the government should do.

Nonetheless, even a cursory review of the examples quoted above from Jack Jennings and Chris Cross—and more could be cited from other papers—will make it clear that just counting something (e.g., charter schools, number of students who move, states with choice programs) will be of limited value. A case in point is that, as Dominic Brewer and Cathy Stasz remind us, in tracking reform, change takes time and it is problematic just finding appropriate words for survey items that adequately describe what should be taking place. Even so, they note, surveys can only provide a snapshot of what is happening at one point in time.

Many of the specific suggestions for NCES data in these commissioned papers call for information about conditions in education that need measurement, such as instructional processes, curricular offerings and content, and the act of teaching. Where the intent of that measurement is to obtain information of a qualitative character—such as describing whether student content standards are equivalent, say, to those of NCTM in math or of NRC in science, or describing effects of instructional changes that can be associated with new student standards, and especially evaluating “educational treatments”—more analytical tools would be required and more micro-level data and case studies that are not feasible for national statistical programs would be needed. The Brewer and Stasz paper reminds us that theories should drive such data collections.

The need to ground data in compelling research theories and findings makes links with research crucial in any field of government statistics activity. I have recently chaired a review of the Joint Program on Survey Methods, an NSF-funded project housed at the University of Maryland, to train federal statistics staff. One of our panelists, an economist, insisted that the training of government statistics staff could be adequate only when statistical training was informed by the linking of statistical planning, design, data collection and analysis with the theories, constructs, and measures developed in academic disciplines related to government functions—whether they be in education, housing, transportation, health, environment, or energy. It was a hard sell and, finally, two of our eight panelists declined to be parties to this advice.

In truth, NCES has been diligent about making these sorts of connections. The message of the authors is that those efforts must continue. One lesson here for NCES is to work out an understanding about what the relative contributions of research and statistics can be, then to call on

other OERI resources when that is appropriate and use the methodologies of both if that is called for. A dazzling example of the using both is described in Jim Stigler's paper "Large-Scale Surveys for the Study of Classroom Processes."

Another lesson for the Center is systematically to search out, use, and adapt research findings in its statistical activity. A fresh example here is the result of 5 years' work on school restructuring conducted by the University of Wisconsin, which was reported a few weeks ago at a seminar in the House of Representatives Rayburn Building. The researchers reported that, for the first time, their work showed a particular approach to teaching was associated with increased student performance—and that increase was for all students. The approach, which the Wisconsin Research and Development Center calls "authentic pedagogy," requires students to think, to develop in-depth understanding, and to apply academic learning to important, realistic problems. I cite this as just one example where the Center should, if these impressive sounding findings hold up, make use of research results as it designs data collections, questionnaires, supplementary information sources, and analyses on teaching.

ON TECHNOLOGY AND ADMINISTRATIVE RECORDS

One implication of the technology and administrative records concepts advanced in these commissioned papers is that the Center would actually collect less data. The emphasis here is on the word "collect." As I read what Fritz Scheuren and Glynn Ligon have said, and others have touched on these themes as well, the combination of electronic record keeping with administrative records is likely to result in much more information contained in those systems than has been the case with paper files. Thus, information that has, up to now, been obtained by surveys may in some cases be available with little investment of either time or funds. At a time when the government funding outlook is lean, while the data demands seem to grow, this combination appears to offer a vision of more for less that is pretty attractive under the circumstances. It does hold out the intriguing possibility that the Center might actually be able to achieve some of the other quality data, analytic, and methodological enhancements called for in this group of papers.

I have two caveats about such a development. First, while NCES has already worked hard on technology advances, it has not absorbed into its data planning and gathering how its own role may need modification in relation to that of states, colleges, and universities and other data providers and users. Perhaps furthest along in this regard are the Center's efforts on library data where self-editing electronic reporting has been developing over several years. Center staff would need to be strongly oriented to the needs of the institutions that are developing electronic systems but, at the same time, alert to how cross-state and national needs could be achieved as well. When and how to intervene, how to play the broker role, and when to subsidize design or planning efforts that can benefit many data needs are examples of the sorts of roles NCES would perform more frequently. Still, these roles are not the same as those exercised in designing another federal survey and may imply a need for staff development activities.

The second caveat is not to leap too quickly to a conclusion that all the public's data expectations for the Center can be derived from electronic systems with their low marginal costs. There will be issues or topics these systems simply do not cover that are a necessary part of an

adequate education information system for the public. Examples include teacher practices and attitudes, student experiences, family circumstances, and attitudes that will and still only be available from individuals. Another example is that the present state-of-the-art in test equating will not permit ready comparison of student achievement results obtained from electronic records and based on widely differing assessment systems. And still another, if the appropriate data to answer a question must come from observations of classroom activity—as in the international classroom video research, which as David Mandel insists is essential when teacher quality is to be adequately measured—electronic administrative records will not contain those either. There will still be much for a statistical agency to do even in this wondrous new electronic world.

ON ANALYSIS

The final matter I want to address in these introductory observations has to do with implications in the commissioned papers that NCES would be more aggressive in making use of data in ways that will better inform the American public about education. This means more analysis. Among other things, it means more tapping of data from a variety of sources—OERI-sponsored research, the Bureau of the Census, systematic reports from organizations such as the Education Commission of the States or the Council of Chief State School Officers, and in-depth studies from individual states or districts, as well as from the Center's own collections. Here, too, the Center can expand its efforts with such organizations to design studies so that their data can be linked and, thereby, made much more powerful for analytic purposes.

Most challenging, however, will be the consideration of breadth versus depth trade-off questions. Reading through the full set of commissioned papers conveys both of these dimensions and could easily lead to numbing paralysis in formulating an appropriate Center response. Program and policy evaluations, longitudinal studies, data at the classroom level, experimental designs, international comparisons, and other costly steps are recommended by the authors.

What to do? The Center has the advantage of context—its place, function, and visibility—in picking the issues about which it will inform Americans. But the cumulative advice in these papers is that to inform well, the Center must narrow the questions it designs statistical activities to answer in favor of providing more powerful and complete data about the questions it chooses to address. The familiar dilemma in the Center is that of the longitudinal studies, such as High School and Beyond and the National Educational Longitudinal Study, where compromises must be made on education questions in relation to, say, employment or family background. Unfortunately, sometimes the result is that important policy issues and data relationships simply cannot be examined with these data. I do not mean to suggest there is an easy answer here, but increasingly the Center is being advised that what constituted adequate data in the past is frequently insufficient for the more sophisticated education information questions being posed as the Center prepares for the next century.

CONCLUSION

These are only a few observations that might be made, and I hope they will encourage readers to delve into the full volume of papers and commentary. The Center certainly received its money's worth from these papers. They are not full of impractical ideas. Instead, they provide the basis for a new vision of the National Center for Education Statistics in a new era.

THIS PAGE INTENTIONALLY LEFT BLANK

2

Tracking Education Reform: Implications for Collecting National Data Through 2010

THIS PAGE INTENTIONALLY LEFT BLANK

Tracking Education Reform: What Type of National Data Should Be Collected Through 2010?

**John F. Jennings
Diane Stark**

ABSTRACT

This paper will focus on the types of data collection the National Center for Education Statistics (NCES) could undertake that would be most useful to policymakers as they address issues of school reform. In particular, this paper will address three questions: 1) what type of education reform data should be collected; 2) how should the data be reported and packaged so that it is useful to policymakers; and 3) what should be done with the data so that policymakers have a better understanding of education reform?

We make three major recommendations to NCES for making its data collection efforts more useful to policymakers. First, we propose that NCES broadly collect data on specific education reform efforts being undertaken at the local, state, and national levels. However, understanding the cost implications of such a broad data collection, we urge NCES to augment its own efforts by compiling education reform data that has been collected by others, and to report both the NCES data and the compiled statistics. Second, we recommend that the education reform data be reported by state, and that information be available to policymakers via the Internet. Finally, we urge NCES to make the data widely available so that researchers and others will be able to undertake in-depth analyses, thereby enabling policymakers and others to have a better understanding of the reform being examined.

WHAT TYPE OF EDUCATION REFORM DATA SHOULD BE COLLECTED BY NCES?

For well over a decade, the nation has been concerned with improving education, especially at the elementary and secondary school levels. Therefore, in order to keep policymakers and the public adequately informed, there needs to be more information collected on education reform. In order to do this, NCES will first need to determine exactly what “education reform” is, and will then need to develop common definitions of the various reforms so that reporting will be uniform. Policymakers and the public will also want to know what effect the reforms have had on student achievement. In making determinations about student achievement, NCES will need to utilize the

National Assessment of Educational Progress and other assessments and studies. Finally, NCES must report on both the statistical data as well as the student achievement information.

Defining Education Reform

Because education is an activity that falls mainly in the public domain, information on education is vital. Voters need information on their schools to determine whether or not to support increased funding for the school system. Similarly, parents need information to determine if the school is providing an adequate education for their children. Finally, educators and policymakers need information to make decisions regarding curricula and approaches to teaching. The National Center for Education Statistics (and its predecessor agencies) has carried out its mandate to report on the status and progress of education in the United States for nearly 130 years, providing policymakers and educators with a broad array of statistics and other data on our schools. As the Center endeavors to collect data on education reform, we would urge a similar broad-based approach.

At first, defining the term “education reform” appears to be uncomplicated because it could be defined simply as any effort undertaken at the local, state, or national level to improve student achievement. However, the task gets more difficult if one considers that what one person may view as a “reform,” another may view as an impediment to improvement. For example, one policymaker may consider private school vouchers as an education reform, while another may view such vouchers as a step toward the destruction of public schools. Similarly, a group of policymakers may advocate opportunity-to-learn standards as an education reform, while others see those standards as requiring unnecessary expenditures in education. In collecting data on education reform, care must be taken to ensure that the Center's data collection efforts in the area of education reform are not viewed as politically motivated or as biased toward one set of reforms.

We believe that only by employing the broadest definition possible of “education reform” and then collecting and compiling data on all reforms can NCES remain an impartial statistical agency. In order to ensure its impartiality, the Center may want to consider convening an advisory committee, made up of individuals with widely varying views on education reform, to guide it as it embarks on this work. Such a committee would not only help the Center determine the broad array of education reforms to study, but also, given the reality of funding constraints, could aid in establishing priorities for NCES's data gathering in this area.

Once the Center determines which education reforms to study, we propose that NCES collect and compile a broad array of data that could be issued in a comprehensive annual report. Such a report would be a vital source of information to policymakers: by having access to information on what other communities or states are doing to improve education, policymakers would be able to “borrow” ideas and secure expertise that would help them in their efforts to improve education in their communities. To our knowledge, no such “encyclopedia of education reform” exists.

NCES will need to broaden its current statistics and data gathering efforts in order to report on specific education reforms. The advisory committee convened to identify the broad range of

reforms could also develop common definitions of terms. While many states and communities appear to be undertaking similar education reforms, such as charter schools or school choice, there are likely to be great and small variances among them. Common definitions will allow for uniform, comparable reporting of data. Once common definitions are developed, NCES should then direct its efforts to collect data that will paint a complete picture of the types of reforms being undertaken, and that will enable researchers and others to undertake analyses that will help to determine which reforms have been successful in improving student achievement and other outcomes.

Many NCES ongoing data collection efforts, studies, and assessments will provide additional information that is needed to determine the effects of education reform. For example, through the National Assessment of Educational Progress (NAEP) and the various NCES longitudinal studies, the Center is able to provide data on student achievement as well as in-depth information about students and the education they receive. This information, along with other data collected by the Center, such as dropout rates and student course-taking patterns, will be helpful in giving policymakers and others the information they need regarding education.

If NCES were to develop a plan for gathering this ambitious set of data on education reform, perhaps there would be interest in the Congress to devote additional funding for the plan's implementation. However, given the current fiscal climate on Capitol Hill, NCES may want to consider other means of financing the data collection. Perhaps the Center could establish a “fund for education reform statistics” and work with the business community and charitable foundations to contribute additional dollars so that adequate funds would be available to carry out this work.

If the Congress does not provide extra appropriations for education reform data collection, and if NCES opts not to create a special fund, then the Center must set priorities for which information it intends to collect directly (perhaps following the advice of the advisory committee). As a means of lessening the burden and allowing for the reporting of a broad scope of education reform data, NCES should also consider becoming a “repository” of data collected by other organizations. Many organizations such as the Council of Chief State School Officers (CCSSO), the Education Commission of the States (ECS), the National Conference of State Legislatures (NCSL), and the American Federation of Teachers (AFT) have been collecting data on various state education reforms. NCES need not “reinvent the wheel”; the information gathered by these organizations could be included in any reports on education reform. We understand that the compiling of data from other sources may raise questions of validity and reliability; but with limited funds available to NCES for a comprehensive education reform data collection effort, it may be the only alternative. NCES could work with CCSSO, ECS, NCSL, AFT, individual states, and private research groups to use common definitions and reporting cycles so that this data could be included in a comprehensive annual NCES report on school reform.

We also recommend that NCES go “on-line” with a page on the Internet devoted entirely to education reform. NCES's education reform “home page” could provide a monthly update of the data that appeared in the annual report on education reform and could also include recent research findings or data from outside groups. For example, there could be a section on charter schools that would include any updated information that may be issued by the Humphrey Institute in Minnesota.

Similarly, any data on state standards-based reform efforts that are issued by NCSL, ECS, AFT, or other groups could be included in the section on standards-based reform.

To assist NCES in thinking about the wide range of school reforms that could be included in its reports, the following sections will briefly describe several major types of reforms.

Standards-Based Education Reform

One of the most prevalent reforms currently being undertaken by states and school systems is standards-based education reform. The underlying premise of this type of education reform is that students, teachers, and parents should know in advance what students are expected to know and be able to do at different grade levels. States develop content standards outlining what is expected in subject matters such as math and science. States also establish performance standards that explain how well a student should perform in a given curricular area, and develop assessments aligned with the state content standards in order to chart student achievement. According to a recent report issued by the American Federation of Teachers entitled *Making Standards Matter*, as of July 1995, 49 states and the District of Columbia were engaged in standards-based education reform. The federal government, through the Goals 2000: Educate America Act and the Title I program, is assisting states and localities as they develop standards and implement this comprehensive reform.

Several national organizations have issued reports on various aspects of state actions in standards-based reform, and other offices in the U.S. Department of Education, such as the Goals 2000 office, have also compiled information in this area. NCES could use the information compiled by these organizations and by other offices in the Department as a basis for issuing a comprehensive up-to-date report on standards-based reform. It would be of great value to national and state policymakers to have a regularly updated status report of which states are developing, have adopted, or are implementing standards in a given subject matter area. A policymaker from a state with no science standards, for example, could use this information to seek out officials from other states that have adopted science standards, and could consult with them about their standards and the process by which the standards were developed. Such a status report would also help state officials to determine where they “measure up” compared to other states implementing standards-based reform. National policymakers would also find the state-by-state standards status report useful as they make decisions regarding the Goals 2000 program and the Title I program.

An integral part of standards-based reform is assessing student achievement. Most states are just beginning to develop assessments based on their standards, and there is little information available to states to help them during this process. Data should be collected on where states are in the process of developing assessments that are aligned with standards, and the type of assessments being developed by the states.

Finally, a controversial component of standards-based reform that some consider key are opportunity-to-learn standards. These standards outline what tangible elements need to exist in a school in order to give a student an opportunity to learn the state standards. They are controversial because they primarily affect “inputs” into the school system (thereby potentially requiring an outlay of funds), whereas the content and performance standards are concerned only with outcomes. For

example, if a state science standard requires that by the 5th grade students should know how to operate a telescope, the opportunity-to-learn standards could require that elementary schools have telescopes. NCES should collect data on the number of states that have developed or implemented opportunity-to-learn standards, and where possible, determine if the implementation of these standards has had a fiscal impact.

Other Reforms

There are several other reforms that are being tried around the country that may affect only individual schools rather than the entire school system, as is envisioned in standards-based education reform. These reforms are either being implemented or considered in nearly every state in the nation and should be examined since they are so pervasive. State-by-state information on the number of schools implementing a given reform would be helpful to policymakers, especially at the national level where some federal programs have been created to encourage certain types of reform such as charter schools. These reforms should be followed over time in order to develop trend data. The following sections outline some of the education reforms that states and communities are undertaking that NCES may want to study.

State Course-Taking Requirements

During the 1980s, after the issuance of *A Nation At Risk*, many states changed state curriculum requirements. Under this reform, students were required to complete more hours of instruction in core academic subjects in order to receive a high school diploma. NCES has done much work in this area through the High School Transcript Studies and should include this updated information in any comprehensive reports on education reform.

Charter Schools

Charter schools are public schools that are by state law exempt from significant state and local requirements. In exchange for this increased flexibility, the schools are held accountable for increased student achievement. It is believed that by exempting these public schools from most rules and regulations, charter schools create an environment where innovation can thrive. They are often created by teachers, parents, or groups in the community. Basic data should be gathered on the number of charter schools in the United States, and because charter schools exemplify the trend in public education of “greater flexibility in exchange for greater accountability,” they should be studied over time to determine how successful the charter school approach is in raising student achievement.

Reform Networks

Hundreds of schools across the nation are engaging in reform activities that employ special strategies for educating children, especially those from disadvantaged backgrounds. Examples of these reforms include Robert Slavin's Success for All, Hank Levin's Accelerated Schools, and Ted Sizer's Coalition of Essential Schools. In an effort to provide support for schools undertaking these

above reforms, networks are created where individual schools can go to get the assistance they need. Data on the number of schools participating in the reform networks by state and, where possible, by school district, should be included in any education reform report.

School Choice

School choice programs allow parents and students to determine which schools to attend. Many states and school districts have choice programs that allow students, within certain restrictions, to enroll in the public school of their choosing; some states also have choice programs that allow students to attend private schools using public funds. Other forms of school choice would include magnet schools created to promote integration of different racial groups. Information needs to be gathered on the number of states and school districts with school choice programs; the number of students participating in the choice program; and examples of state or school district choice policies (i.e., open enrollment, student selection criteria, and so on).

Private, For-Profit Companies

Several school systems have contracted with private, for-profit companies to run their school systems or to operate individual schools. Data should be gathered on the number of school systems that have made such arrangements, which companies have been involved, and general characteristics of the schools or school systems that are affected.

State Takeovers

While not quite an education reform in the traditional sense, in an effort to boost student achievement, some states have taken over failing school systems. Data should be gathered on the number of states that have a policy or law allowing state takeovers; the “triggering” conditions for such state intervention; and the number of school districts or schools that are affected by state takeovers.

School Finance

As part of any report on education reform, data should be included on state and local efforts regarding the financing of education. It appears that it is a trend among several states to limit or curtail the public financing of education through local property taxes and to instead fund schools through state taxes. Meanwhile, there are several court cases pending that call into question the disparate per-pupil expenditures existing within states. Further, there is some concern about differing per-pupil expenditures within school districts. Data should be reported, on a state-by-state basis, on all aspects of the school finance issue. On the national level, if the dramatic funding cuts that have been proposed for federal elementary and secondary education programs become law, it would be essential to know how these reductions affect school finance at the state and local levels, including the number of students who no longer have access to the programs supported with federal funds.

Block Grants

There are various proposals before the Congress to create education block grants. Block grants are formed by taking a number of separate programs that have similar purposes and combining them into one large program with one set of requirements. Block grants usually mean increased flexibility in the use of federal funds by states and school districts, but the creation of a block grant is also usually accompanied by decreased federal funding. If such education block grants are enacted into law, they will have a considerable effect on the financing of education, and should be studied to determine their impact. Several states are also “block-granting” state categorical programs, and these efforts should be reported.

School Infrastructure

Another element of school reform that is not instructional-based is the renovation or rebuilding of aging school buildings. As illustrated in Jonathan Kozol's *Savage Inequalities*, some school buildings are in such poor condition that they are literally falling apart, lack working plumbing, and are unfit for human occupancy. In other instances, the buildings lack the necessary facilities (such as science laboratories) to adequately provide the type of education that is needed in today's high-tech world. Data on aging school buildings and the state of school facilities need to be included in any reports on education reform so that state and national policymakers will have information about the general condition of schools across the nation as they make policy decisions.

School-to-Career Reforms

Through its Data on Vocational Education (DOVE) system and the first and second National Assessment of Vocational Education (NAVE) reports, NCES and the Office of Educational Research and Improvement have gathered and reported significant data on the status of vocational education in the United States. Continuing these data collection efforts is extremely important as reforms in vocational education, such as “tech-prep” programs and school-to-work transition initiatives, grow in popularity among the states, and as these initiatives become a primary instrument for reforming secondary schools.

Home Schools

Home schooling is a growing trend resulting from parental desire to oversee all aspects of a child's education. Reasons cited for home schooling range from religious beliefs to dissatisfaction with the education provided by traditional schooling. It is important that data be collected on home schooling in order to get a clear picture of all the endeavors being undertaken to educate children. Again, state-by-state data on home schools are essential, as well as a description of the oversight governance of home schools in each state or locality. (That is, does the state, school district, or other entity ensure, through assessments or other means, that home-schooled children are receiving an adequate education?)

Postsecondary Education Reform

Nearly all the above-mentioned reforms affect only elementary and secondary education. Several states are beginning to examine their postsecondary education systems and are considering implementing reforms to improve teaching in higher education. Because this is an emerging reform, data on the number of states that have postsecondary education reform initiatives as well as information on the content of these initiatives would be very useful to policymakers at all levels, especially as more states embark on higher education reform.

Assessing the Goal of Reform

Elementary and Secondary Student Achievement

In studying all these reforms, NCES should gather data on what is the intended outcome of the reform. For example, nearly every reform mentioned above would probably have as one of its goals increased student achievement. NCES, through its traditional measures such as the National Assessment of Educational Progress (NAEP) and through other surveys and studies, could measure the impact of various reforms on student achievement. In particular, as NAEP becomes more aligned with the national academic standards, it can be an important vehicle for measuring student achievement under standards-based education reform. Other measures, such as the National Education Longitudinal Study and the High School and Beyond reports should also be used to determine if various reforms have improved student achievement. Additionally, NCES should also consider other sources of information on student achievement, such as state assessments and college entrance examinations. Each of these assessments is potentially a rich source of information on student achievement along with information on student characteristics.

Postsecondary Student Achievement

NCES should also look into measuring the achievement of postsecondary students. As was mentioned above, attention is beginning to turn to postsecondary education reform, with talk of developing academic standards similar to those developed for elementary and secondary education. Currently, NCES collects data on such factors as postsecondary enrollments, completion rates, the number of students receiving financial aid, and faculty and institutional characteristics. The Center does not assess postsecondary student achievement. While current budgetary constraints may preclude it, we would suggest that NCES consider either expanding NAEP to include students in postsecondary education or to develop a separate NAEP-like assessment to chart the achievement of students who have continued their education beyond high school.

Other Goals

NCES should also collect data that will enable policymakers and the public to determine if other goals of the education reform initiatives are being achieved. For example, while increased student achievement may be a goal of a tech-prep program or a school-to-work transition program, those reforms also have as their goal developing student occupational and work skills. Similarly, school choice programs may not have increased student achievement as their primary outcome;

rather such initiatives may be solely designed to give parents more educational options for their children. These other goals of education reform should not be overlooked by NCES.

International Comparisons

In recent years, there has been increased demand among policymakers and educators for information on how student academic performance in the United States compares with students in other nations. NCES, through the OECD International Education Indicators Project and other international studies, has helped to provide that needed information. In the future, individual states will want to know how their students compare to students in other nations. NCES should be prepared to provide such comparisons.

PACKAGING AND REPORTING OF DATA

In order to be most effective in reaching policymakers, NCES needs to pay particular attention not only to the scope of data collected but also to its packaging and reporting. On all levels of government, policymakers' attention is being drawn in several different directions on many issues. Therefore, concise, timely information that is relevant and easily accessible is essential.

Readability and Relevance

Information for policymakers should be in a form that is easily understood since they are often dealing with several divergent issues at once and may not be experts in education. Reports that contain executive summaries as well as charts and graphs with easily understood explanations best suit the needs of policymakers. This sort of concise information would also be of use to a wider audience such as parents, the public, and the media.

Further, in order to meet the needs of policymakers, reports issued by NCES should contain data that is regionally or locally relevant. We recommend that NCES consider studying education reform on a state-by-state basis, and issuing annual reports on state activities. The state-by-state information would assist policymakers at all levels to understand the impact of education reform. More detailed information could be put “on-line” and be available to state legislators and others who may need more in-depth data on the condition of education reform within an individual state. Also, NCES may want to create a special state education reform hotline so that policymakers and others could have immediate access to the information.

In compiling the state-by-state data, NCES should, to the extent possible, collect data on education reform at the sub-state level, especially at the school district level. While we understand the cost and data reliability issues involved in sub-state reporting, we believe that this information is essential in helping parents and local communities understand what their schools are doing and if the reform being implemented is effective. This, in turn, helps policymakers do a better job at representing their constituents' views on education, as well as provides them with information they need to make informed decisions. If the Center is able to collect data on school district education reform, these data should include information on the demographic and economic makeup of school

districts so as to give the reader a context for the information. Also, as a means of reducing costs, NCES could draw on state assessment data and other state reports in its sub-state data gathering efforts.

Timeliness

One of the ways that NCES can be most effective in meeting the needs of policymakers is to anticipate when certain data will be needed. This is especially true in the legislative arena. For instance, if the Congress is debating a bill designed to reduce student dropout rates, a report filled with state and local dropout data and an analysis of state and local dropout prevention programs should be issued before that the debate occurs so that informed decisions can be made. Usually, federal education programs are authorized for a set number of years, and near the end of that authorization period, the Congress begins to consider the effectiveness of the program and whether it should be continued. To the extent possible, NCES should pay close attention to the reauthorization schedule, and time the issuance of reports to coincide with that schedule.

Also, it goes without saying that NCES should provide up-to-date information whenever possible. Policymakers need to have the most current information available so that they can make decisions based on what is happening in the present, not on 5-year-old data. By updating the education reform data that would appear on NCES's "home page" on the Internet, policymakers and others would have immediate access to the most recent information available on education reform.

ANALYSIS

The most important aspect of making NCES education reform data useful to policymakers is to have more analysis of the data. While policymakers find it useful to know the postsecondary attendance rates of students in their school district or to know how well their students did on the state NAEP assessment, they are more interested in knowing why a certain achievement trend is occurring. In education reform, it is essential to know possible reasons why one reform succeeded in a state while another one failed so that policymakers can fine-tune programs or make other necessary adjustments. This information can be gleaned only from analysis of data that have been collected over time.

While NCES does engage in some data analysis, far more needs to be done, especially if the Center begins an effort to study education reform. NCES may not have the capacity at this point to conduct the kind of analysis that is necessary, nor may it be an appropriate function of the Center. However, it is our hope that the information collected and compiled by NCES on education reform and student achievement will spark the interest of researchers to look deeper into the data. In order to move that process along, we recommend that NCES, together with the Office of Educational Research and Improvement (OERI), convene a conference to discuss how NCES and OERI can make data more accessible and promote analysis.

To the extent that funding is available, NCES may also want to explore the option of contracting for such analysis. If NCES decides to promote analysis of education reform data through

contracts, it should do so with several target audiences in mind. For the needs of policymakers (and for that matter, the media as well as the general public), the analysis and reporting should concisely explain the effect of an education reform and give possible reasons why this effect occurred. For the needs of educators and researchers, the analysis should be more complex, painting a more complete picture of how an education reform affected student achievement. Policymakers may also return to the more in-depth analysis as situations warrant.

SUMMARY

In conclusion, in order to meet the needs of policymakers, we recommend that NCES expand its data-gathering efforts to include reporting on specific education reform initiatives being implemented at the local, state, and national levels. Once the data are gathered, NCES should package the information in an annual report that would include education reform data collected by other organizations. The information contained in the annual report should also be available through the Internet and be updated as newer data become available. Summary information should be provided to policymakers, and NCES should try to issue the release of information in a manner that provides policymakers with timely, needed information as they begin to debate an issue. Finally, NCES needs to encourage more analysis of education reform data by researchers so that policymakers and others will be better able to understand possible reasons why a certain education reform succeeded in improving student achievement or why it failed.

This paper has outlined a rather ambitious set of recommendations for NCES with regard to policymakers' need for education reform data and statistics. We realize that meeting all of our recommendations may be impossible, especially given today's funding realities on Capitol Hill. However, we believed that we had to set out a broad vision so that actions could be taken to achieve it, if only partially. NCES must be sensitive to the changing needs of policymakers not only to be able to serve them better but also to remain a viable federal agency.

Where Are We Going? Policy Implications for Data Collection Through 2010

Christopher T. Cross
Amy Rukea Stempel

INTRODUCTION

We know that education reform is happening and that the academic achievement of American students is lagging behind what is expected of them both in our nation and in the world. However, we do not know what links education reform efforts to changes in academic achievement. The collection of educational statistics by the National Center for Education Statistics (NCES) can assist attempts to pinpoint the relationship between reform and achievement. By isolating different aspects of education reform and attempting to remove superfluous influences, we can begin to cull out those changes that are transforming American schools from those that are not.

What is rapidly becoming apparent is that even though we have some idea as to what affects academic achievement, we are still floundering in our efforts to reform the nation's education system. There seems to be no one reform that actually accomplishes all we need it to do, as much as proponents of various reform agendas might wish it to be so. However, there may be clusters of reforms that, when integrated and advocated with intelligence and moderation, might actually produce results. Unfortunately, we have little data to support reform recommendations of this type.

Besides the statistically reported national academic achievement and the change in that achievement over time, the Council for Basic Education (CBE) would like to suggest collecting data on other, less immediately apparent, factors in education reform. We do not discount the necessity of gathering statistical information on achievement; however, we believe that our biggest pitfall in education has been ignoring the more subtle issues affecting reform.

Beyond data collection, NCES might consider devoting more attention to analysis of that data. If in-depth analysis is not feasible, providing readers with possible considerations for analysis would be helpful. There is no doubt that the efforts of NCES to encourage wider use of their data have met with success. CBE would like to see this effort continue to be a NCES priority. The challenge of statistical analysis is to get to the heart of why and how education reform helps or hinders educational achievement and student learning. To do this, we believe that the questions asked will have to be modified to capture the inherent ambiguity and interconnectedness of the educational endeavor.

One of our major concerns is the validity of the data collected. Over the years, we have discovered that self-reported data are notoriously unreliable. For example, teachers over-report their implementation of reform; principals assure the public that their schools are consistently performing better now that “X” and “Y” reforms have been mandated; and parents seldom understand reform agendas enough to make informed decisions about the truth or falsehood of these statements. Without reliable information, we run the risk of making ill-informed decisions that will do more harm than good. By ensuring that the data collected are valid and by encouraging analysis, if not providing it, the information collected can be put to direct use by policymakers, educators, parents, and students.

VALUE ADDED

Information on achievement is a “slippery fish to catch.” We have learned over the years that there are many factors that affect student learning that are not school functions. For example, countless statistical surveys have shown that socioeconomic status, specifically the mother's educational level, is more of an indicator of student achievement than any other factor. With all these secondary indicators of success floating around, we need to be careful to what we ascribe achievement and how we report it.

Therefore, we believe strongly in the need to examine the value-added issue more closely. Given a variety of starting points, what does a particular reform effort or combination of reform efforts add (or not) to current student achievement in a particular school or type of school? If we can begin to address this question, we will be well on our way to establishing the relative merits of various reforms. Data collection in this instance should focus on the school, the types of interventions offered, and how they affect a variety of students attending. While other factors may be stronger indicators of success, schools are ultimately more manipulable than an individual's SES or parental education levels.

There are many reasons to worry about education in America. Our highest achieving students are lagging behind their world counterparts. Even more debilitating are our inner city schools, many of which seem barely able to teach students to read, let alone succeed in the world. Scores on achievement tests alone may or may not validate the success or failure of reform efforts for both the highest and lowest achieving students. However, looking closely at the changes in achievement scores and their relationship to many indicators, rather than accepting them at face value, would come closer to answering the question: What is the value added by this school? this reform? this program? Fundamentally, we believe that this is what people want to know.

We have developed our recommendations with our need for reliable information and our belief in the need to articulate and ferret out the value added by various reforms in mind. Education is complex and educational data collection needs to reflect this. Unfortunately, this presents a different set of implications to those collecting data. The point of this paper, as we understand it, is not to support the status quo, but to go beyond the traditional role of NCES and challenge it to find solutions to the intricacies of reliable, complete, and insightful educational data collection.

The reform agendas discussed in this paper are based on an assessment of what is present on the national reform horizon. Subsequently, we will be discussing teacher education and development; issues of school governance and organization, such as site-based management, home schooling, charter schools, magnet schools, and voucher programs; articulation between levels of schooling; educational technology; academic standards; and assessment. As you will note, there are particular dimensions of these issues in which we are primarily interested that will constitute the bulk of this paper. We will also discuss the costs and benefits of some of our suggestions, followed by a brief discussion of the implications these suggestions have on data collection methods.

THE DIMENSIONS OF REFORM

Systemic information about education reform efforts is crucial because the long-term health of our national education system requires that we radically change how we educate our students. How does our education system work? What are its flaws? Why is change so difficult? By examining and collecting data on crucial systemic points, we can begin to address the complexity and interconnectedness of education reform. We are primarily interested in how these reforms are being used (or not) in the system. However, more important is what effect these reforms have on academic achievement. Throughout this paper, we will consistently return to these two issues.

Teacher education has been under siege for many years. Tomes have been written on what teachers are not required to do to become certified and how undertrained and ill-used they are. CBE suggests an attempt to evaluate and measure the quality of teacher training, professional development, and professional support in an integrated way. Rather than providing a catalogue of courses required to become a teacher, it would be more informative to provide information about the philosophies of particular schools of education and how these philosophies are carried out both in the teacher training curriculum and later in the K–12 curriculum.

For example, how does the curriculum of teacher training institutions that advocate student-centered learning reflect that philosophy? How is a vision of the educational experience linked to its practice? How do the professors responsible for training future teachers conduct their classes—lecture, group activities, socratic seminar? How do teachers trained in a particular institution translate their training into the classroom? Discovering what actually goes on in teacher training programs and classrooms, rather than what is reported to happen, would help us make decisions as to what works and what does not.

Another reform agenda, “site-based management,” has recently been coming under fire from the public and policymakers. We would like to determine the extent to which site-based management exists and is working nationwide. How is site-based management defined? In schools where it is said to exist, how is it implemented, and are the results substantially different in practice from the norm of top-down management? What is gained and what is lost by switching to site-based management? Are there similarities in site-based management and the administration of private schools that are worth exploring? And last but not least, does successfully implemented site-based management have a positive effect on student learning?

Given that the purposes of schooling are to prepare students for a personally and professionally productive life, more consideration needs to be given to how the system works together (or not) to support a consistent purpose and vision of education. One way to begin this, CBE believes, is to develop a national survey that examines the articulation between the levels of schooling. Anecdotal information indicates that it is surprisingly uncommon for teachers or principals to examine what happens in the levels of schooling before or after the level that is their immediate responsibility—whether it be 7th grade specifically (teachers) or middle school in general (principals). Similarly, we suspect it is uncommon for teachers or principals to examine what teachers of other disciplines at the same level do and expect in their classroom. We would like to know if there is horizontal and vertical coordination of curriculum, and what are the expectations within individual schools and throughout school districts. For example, what are the expectations of the elementary schools in a district? Are the middle schools in the district aware of those expectations and do they begin where students have left off? Does a teacher in 8th grade, for example, know what other 8th-grade teachers do and expect in their classrooms?

There are other issues currently in the public interest that are not central to student instruction but that deserve exploration: for example, school safety, moral and character education, and the effects of parental involvement on students' educational achievement are “hot” issues in the reform debate today. What these proposed reforms have in common is that they do affect student achievement, but no one is sure how they interact with other elements of reform. Does a rigorous character education program improve academic achievement or school safety? Does simply making a school safe improve behavior or academic achievement? Is parental involvement in their children's education linked to the other factors discussed? An informative survey would try to tease out the different strands of reform and establish how they interact. No doubt this would be a complex data collection to design; but one thing is certain: it would be incredibly helpful to all involved in education.

Data needs to be collected about several more recent phenomena: technology, home schooling, charter schools, magnet schools, and voucher programs. How do these efforts at school governance and organization affect academic achievement? Technology offers us a new way to gather and disseminate information and provides an ease in data, word, and information processing previously unknown. Unfortunately, the education community, for lack of money and political power, is far behind the technological boom. In order to truly document how technology affects learning, we will need to document the uses and abuses of technology in classrooms across the country. For example, recent newspaper reports have revealed a frustration that technology is not the panacea it was first touted to be. We doubt that there is a single panacea, but even so when the reports were followed up, it was discovered that the computers were being used as high-tech workbooks and that the learning process had not significantly changed. Just because schools have access to technology does not mean they use it to its fullest capacity. How schools use the technology they have is one of the more crucial questions of the next few years.

A related issue is the state of the technological infrastructure of the nation's public schools, which is dismal. No one really knows how much money it would take to upgrade them. The true extent of the problem is often obscured by the massive amount of speculation and little hard data. Often surveys ask if a school uses computers in its classrooms, but not how many modems or

Internet connections each classroom has. As the uses of technology grow, so should our interest in how schools are putting their technology to use.

As with the technology issue, home schooling in the United States is rapidly increasing. However, we lack ready information as to who does it, why they do it, the average number of years they do it, the most popular grades to home school, and how technology has affected the home-schooling boom. We also lack information about the academic achievement of these students and how they fare when (or if) they return to school. Where home schooling works well are there lessons that can be transferred to schools about use of time, student/teacher ratios, and options for creativity? It would also be important to find out whether home schooling is a growing option or a passing fad. Although home schooling is often ignored, we believe this segment of education is one of the fastest growing, with research potential yet untapped.

Charter schools, or public schools that are given permission to ignore certain rules and regulations in order to try to increase student achievement, are also growing by leaps and bounds. Because they are free of crippling bureaucracy, charter schools have the flexibility to implement reform decisively. However, for every success story, there are instances of financial abuse and declining test scores. We suggest attempting to evaluate the performance of charter schools nationwide. One fundamental tension in education addressed by charter schools is accountability versus flexibility: what is gained and what is lost when schools, administrators, and teachers are given autonomy in their decision making?

Magnet schools are public schools that have been allowed to choose a particular focus for their academic activities. Perhaps the most famous of these is the public High School for the Performing Arts in Manhattan, which was the inspiration for the movie *Fame*. Aside from the performing arts, schools can be organized around marine studies, the military, technology, or even traditional pedagogy, to name a few. Students must choose, and be chosen by, the schools. The interesting elements of magnet schools are the focus provided by the organizing principle and the element of choice. Does a thematic approach to an academic education provide a focus and motivation for students? Does the fact that students must choose the school and make a commitment to it, as well as be chosen by the school, increase academic achievement?

The increase in voucher programs throughout the country also deserves consideration. School vouchers provide parents with a certain amount of money per student, commensurate with per-pupil expenditures in that district, which they can take to either public or private schools within their community. Is there an increase in academic achievement for students whose parents use vouchers to choose schools? How much of an increase or decrease in achievement is due to the particular school and how much is due to the act of making a choice and a commitment? While difficult questions to answer, the results of such an inquiry would enable parents and policymakers to make decisions about school choice more effectively.

Besides systemic information, there is also information about what occurs in the classroom that is crucial to educators and policymakers. While systemic reform is necessary and desirable, the work of education goes on in the classroom, and it is there that we must look for the bulk of our information. Given that radical systemic reform is still a long way off, a detailed look at what is currently happening in classrooms nationwide will help us in our more immediate future.

The current debate about national standards is an interesting one at the policy level; however, there is no information about the implications of the standards setting projects on education reform. Do academic standards improve achievement? How are academic standards being implemented—top down or bottom up? In states that have developed academic standards, is curriculum being designed with standards in mind? Is professional development provided for those teachers who are expected to implement standards?

We suspect that most who support standards honor them in the breach; the inertia of schools tends to slowly make its way back to the status quo after a vigorous attempt at change. Effective data collection and analysis in all areas of education reform will enable policymakers, educators, teachers, and students to take the pulse of the system and measure what their responsibilities are in order for true reform to occur.

To even begin to measure the effect of academic standards in the schools, we will have to measure the quality and uniformity of assessments and testing in the classroom. CBE has discovered, in the process of doing business, that there are incredible assumptions made by parents, educators, and policymakers about the verity and uniformity of individual student grades. Therefore, we also believe that it is necessary to have some data collected about individual teacher grading schemes such as how teachers determine individual student grades, how they construct their own assessments, and what they fundamentally want their students to know and be able to do. Are teacher grading schemes uniform? This is information the public needs to know.

COST-BENEFIT INFORMATION

Hand in hand with information on instruction and learning, we also need to analyze the costs and benefits of various reform efforts. Reforms that might at first appear to be expensive prove to be quite cost effective when examined from the point of view of the benefits they will provide in teacher training, reduced need for remediation, and student focus, for example. Alternatively, reforms that at first seem inexpensive might prove to cost schools more money if they are not well-organized and meaningful.

For example, what are the initial costs of infusing technology into the schools? What are the maintenance and upgrade costs of educational technology? What services can ports to the Internet provide students and teachers that might take financial and time pressure off school districts—teacher training and development and access to archival records, for example?

Most interesting to reform efforts in the days of shrinking budgets is how successful schools streamline the use of limited resources. Perhaps the most useful question would be how do these schools set their instructional, hence financial, priorities and what are these priorities? Successful schools often employ creative methods to develop resources they believe necessary to instruction and learning. What are these creative methods? How well do they work? Are they personality dependent?

DATA COLLECTION

We understand our role in this process to be that of provocateur; hence, some of our suggestions of areas to explore will require different methods of data collection than those used in the past. We suspect that data collection efforts would have to become more delicate, sensitive, and focused. As always, there are advantages and disadvantages to such a change. By isolating very specific information, there is a limit to the number of ways it can be used. However, if we can determine that the data collected, no matter how focused, are able to provide pertinent information about the reform movement, then the trade-offs would be worth it.

Statistical data collection is certainly useful. However, CBE believes that its use is limited in fields like education where success or failure depends on a host of often conflicting variables and human imperfections for which there exist limited methods to control. We suggest using pure statistics as a tool for analysis, not as an end in themselves. In other words, provide people with the initial information to investigate “why” and encourage them to do so.

One way to do this is through an integrated combination of quantitative and qualitative information working together to answer questions about education reform. Because self-reporting is unreliable, we need to consider other options for data collection such as independent data collection agents or “inspectors” who are responsible for evaluating the relative levels of existence of various reform efforts. We do not let students grade themselves, so it seems equally self-defeating to let those who participate in schools be the sole assessors of their own success or failure. While perhaps blurring the line between data collection and research, we currently see no other way to ensure accurate, reliable information about systemic activity.

These data collections could take the form of both longitudinal studies and single-point studies. For example, in the teacher training example for data collection mentioned earlier, longitudinal studies would be most effective to determine the influence of teacher training on future classrooms, while a single-point study could help determine the context of teacher training.

We also recommend stair-step surveys to link information at the school, district, state, and national levels to what is going on in individual classrooms. Surveys of this sort would inform us as to whether the coordinated reform effort is going well or not.

We understand that we may be recommending an extension or redefinition of some of the activities of NCES. Please view our recommendations in light of our mission and position in the field of education reform. After 39 years of advocating rigorous liberal arts education for all students K–12, we believe that instruction and learning are not simple processes to be easily understood and broken down. As valuable as NCES statistical data are, we believe they can be made more valuable by extending their purpose and offering users even more reliable information, more subtly realized.

Discussant Comments

MARY J. FRASE

These two papers are very similar, and there is considerable overlap in their perspectives and some duplication in specific recommendations. Both represent the perspectives of policymakers and try to outline what the authors feel would be useful information about reform for policymakers. A major difference between the papers is the emphasis in the Jennings and Stempel paper on the need for state-level and sub-state data. Because of the similarity and overlap between the two papers, I will discuss them together, rather than each one separately.

Topics Related to Reform

Out of the two papers one can assemble a long list, a wish list, of reform-related topics one or both mention as being useful to have information about. The former Commissioner, Emerson Elliott, mentioned that planning efforts during his tenure tended to produce similar results—long lists of topics people wanted information about but no suggestions about what to delete. Out of these papers, I came up with 35 separate issues or topics (see Appendix A to this paper), and I probably missed some. In this case, the long list was probably deliberate in light of the charge to the authors not to take fiscal constraints into account. Jennings and Stark wrote that they purposely took a broad view, realizing that NCES probably could not do everything they mentioned.

One issue that needs to be raised at the beginning is what is meant by “reform.” Is it just another word for “change”? The Jennings and Stark paper warns that NCES should take a very broad view of what constitutes reform so the agency would continue to be seen as impartial, i.e., not endorsing one approach or type of reform over another. But does that lead to looking at everything? Is there a trade-off here between depth of information and impartiality? If NCES were to gather a great deal of information on few, high-priority “reforms,” would the agency be seen as endorsing those reforms? If it gathers information on a wide range of reforms, the result may be breadth but not depth of information. Jennings and Stark suggest an advisory panel to help set priorities about which reforms to follow. In the current fiscal situation, NCES cannot expect more money, so someone would have to make some choices. Perhaps the best way to think of these papers is to view them as a menu from which NCES could choose topics.

Toward that end I have tried to think not so much about specific “reforms,” which may have relatively short “half-lives,” but rather types of information about reform that might be useful to have, regardless of what the reforms are. I grouped the topics the two sets of authors mentioned into a limited number of categories in order to see if that might lead to some insights about what NCES might pursue out of this menu. Most of the 35 seem to represent one of four types of information.

- *How much reform is happening?* How many students are being home schooled?

How many school districts have school choice plans allowing choice within or across districts?

How many states are pursuing standards-based reform?

- *How is reform being carried out?*

What oversight mechanisms are in place for home schooling?

What types of school choice plans are being used?

What does “standards-based reform” mean in the various states?

- *What is the effect of reform on achievement and other outcomes?*

How well do home-schooled children perform relative to those in public or private schools?

How is the availability and the utilization of school choice related to student achievement, motivation, and parental satisfaction/involvement?

- *What basic kinds of data are needed to provide contextual or baseline information for reform efforts?*

How can NCES provide contextual information about school finance, student mobility, school facilities, postsecondary achievement, and teacher development?

I will briefly discuss each of these four types of information, as well as their relationship to the NCES data collection program. This is not the only way nor necessarily the best way to group these topics, but the basic point is the need to think systematically about categories of information rather than about specific, relatively narrow topics or issues (i.e., not to miss the forest for the trees). What types of information are most appropriate for NCES to gather directly and what role, if any, might NCES play relative to other kinds of information?

How Much Reform Is Happening?

This is the simplest of the categories. It involves tracking the extent of reform activities by collecting counts of different activities, once decisions are made about what to count. Such information can be collected with fairly simple questionnaires. While NCES can do some of this, it may not be a good use of NCES resources. Other parts of the Office of Educational Research and

Improvement (OERI) and the Department of Education are already doing some of this, as are other organizations such as the Education Commission of the States. The regional laboratories are possible candidates, and Emerson Elliott mentioned that there is already some interest among the labs for doing something of this sort relative to charter schools. Charter schools are one of those relatively rare phenomena that NCES is not good at capturing.¹

The role of NCES relative to this type of information might take two forms. NCES could collect it directly where there are existing vehicles for doing so, which might include the Common Core of Data (CCD), Schools and Staffing Survey (SASS), or Fast Response Survey System (FRSS). Alternatively, NCES could play a brokering role, where the agency would determine what holes existed in terms of missing information and would help identify other ways to gather such data, perhaps involving the Forum and the Cooperative Systems.

How Is Reform Being Carried Out?

Some of this also involves collecting counts of activities, but at a more detailed level. What kind of approach to gathering the information is most appropriate depends on the level one is interested in, i.e., state, school district, school, or classroom. Some kinds of information could be gathered with more detailed questionnaires. Others may require case studies of how reforms are being implemented, since the research literature on implementation reveals there is much slippage between written policy and what happens in the field, in this case, the classroom. The same reform can look very different in different places, and different reforms can end up being implemented in similar fashions. Here there would be a place for qualitative or observational techniques. (It is interesting that nearly everything mentioned in the two papers is either an input or an outcome variable, but there is little mention of the processes linking the inputs, including policies, to the outcomes. That mirrors the strengths of NCES—much progress in developing information and indicators on inputs and outcomes, but relatively weak on measures of process, i.e., what goes on in the black box, in the classroom.)

What Is the Effect of Reform?

This is the toughest and most problematic type of information for NCES to gather. The agency is not in the business of program evaluation. There is an important difference between monitoring what is happening in education and evaluating those happenings. The first is an appropriate role for a federal statistical agency, while the second is not. Furthermore, establishing the impact of a particular program is difficult and complicated, and large-scale national surveys such as those that NCES typically conducts are not well suited to doing such evaluations. The difficulties are illustrated by an NCES publication released 3 years ago.

One provision of the Hawkins-Stafford Amendments of 1988 reauthorizing NCES was a mandate to study the effects of higher standards (as the result of reform) on student enrollment and persistence, academic achievement and graduation rates. In the end, the report consisted primarily of two types of information: an enumeration of the types of reforms raising student standards that had been enacted between 1984 and 1990 and the number of states involved (the first category of information, how much reform is happening); and secondly, a description of trends in student outcomes over the same period. The report made it very clear, however, that one could not link the

two types of information together in a causal fashion. The two sets of events had occurred during the same time period, but one could not conclude that one caused the other. The Executive Summary of the report emphasized that point strongly in the following passages (Medrich et al. 1992, pp. vi and vii):

Even though the states are increasingly active in defining student standards, linkages between these initiatives and student outcomes are difficult to measure for a number of reasons:

- States have adopted different reforms at different times, and no two states have adopted the same exact requirements;
- Even in cases where similar types of reforms can be identified among several states, there is much variation in how these initiatives have been implemented from state to state;
- While some reform activity occurs at the state level, far more occurs at the school district, school, and classroom levels; and
- Over time, demographic shifts have been dramatic in many states, and it is difficult to control for the effects of reform, over time, on different populations.

Although it may be possible to ascertain whether changes in student outcomes have occurred in a positive direction over time, this only suggests that state reforms may be associated with these outcomes. Given the caveats noted above, linkage in a statistical sense cannot be substantiated

. . . [I]n order to establish linkages between state reforms and student outcomes, it will be necessary to examine in more detail the ways in which states implement reforms (the translation from policy to practice) and the extent to which reforms change practice; the impact of specific reforms on local school districts and classrooms; and changes in curriculum content and the quality of instruction associated with, or resulting from, reforms of student standards.

Emerson Elliott has emphasized that there is an important distinction between research and statistics. This is an area far better suited to research—where one can gather pre-reform data, study implementation, gather information about all the contextual factors involved, and look at the “value added” by reform mentioned in the Cross and Stempel paper—than to large-scale data collection.

What Kinds of Contextual or Baseline Data Are Needed?

The last set of topics involves areas where the authors felt data that could serve as important contextual or baseline information for policymakers interested in reform were not available. NCES already collects some of these types of information, but may need to collect more or make the availability of such information more widely known. For example, the longitudinal studies and

NAEP are mentioned in the papers, but not SASS. NAEP was suggested as a way to gather more information on teachers' education and development, but SASS already collects a great deal of such information. The agency may need to do a better job of making people aware of SASS and the type of data it collects. In other cases, NCES is exploring ways of gathering the kinds of data mentioned, most notably, postsecondary assessment data. For some other topics, NCES has considered the need for such data, but is not actively pursuing ways to collect them for a variety of reasons, including cost.

Strategies Related To Data About Reform

The other broad theme in these papers is strategies that NCES might pursue in a variety of areas, including data collection, what to do with the data once they have been collected, and dissemination. The following discusses a few of these that I found interesting.

Jennings and Stark suggested preparation of an annual report on reform, an “encyclopedia of reform,” that would include how much of various kinds of reform was going on, with information by state. The report would include not just NCES-generated information, but also data collected by others. They suggest the information could be put up on the Internet and updated as new information became available, rather than waiting to release all of it simultaneously in the publication. Such a report is an interesting suggestion, but probably would be more appropriate for another organization, such as OERI's Office for Reform Assistance and Dissemination (ORAD) or possibly the Goals Panel (if it survives), which is already doing a considerable amount of this type of activity.

In terms of NCES using data that have been collected by others, which is discussed at length in Fritz Scheuren's paper on administrative record data, I have one concern. NCES needs to be very careful about the caliber of such data that it releases, either electronically or in publications. Utilizing and releasing such data, despite being issued with many caveats, will be seen by the outside world as an endorsement of them by NCES. How this is done, including whether NCES imposes some standards the external data must meet, will affect the likelihood that such a strategy could compromise the reputation of NCES for accuracy, reliability, and impartiality.

The papers also emphasize the need for more analysis of the data NCES collects and production of reports that are more attuned to the audiences of policymakers and the public. NCES is already vigorously pursuing these strategies. The new Education Statistics Services Institute (ESSI) should facilitate doing more and doing it better in these areas and the agency would welcome additional concrete suggestions. The trick is to produce material that is policy-relevant, but does not cross the line into policy evaluation or policy recommendations.

It is also intriguing to think what a paper written by a researcher interested in reform would have looked like. These two papers reflect what policymakers are interested in right now. What is discussed is a lot of separate topics, representing breadth, but not necessarily depth, of information. My guess would be that a researcher addressing this topic would have focused on in-depth analysis of a few topics or on a structure to monitor change, apart from the current “hot topics” and would have come up with suggestions for systematic, in-depth studies, either as new surveys or as components or modifications to existing surveys.² For example, they might have proposed a

longitudinal study of schools as a component of SASS. (Such a survey might be good for gathering three of the four types of information identified in these papers, but not for the effects of reform. It could be used to track what kinds of practices and policies were pursued in a group of schools over time, but to collect information on the effect of those practices would require collecting information on students, with before and after reform measurements, as well as much contextual information.)

The difference between that sort of suggestion and what is in these papers reflects the differing demands of the various audiences of NCES. While everybody wants more information, researchers tend to want in-depth, systematic studies, while policymakers are looking more for breadth rather than depth for issue-brief type of information that is readily accessible, i.e., for sound bites. Part of the challenge for NCES is to provide useful information to both types of audiences in a time of fiscal constraint.

One strategy I feel needs attention, but which was mentioned in neither paper, is how NCES can do a better job of identifying (and gathering information about) new issues that are emerging on the horizon. It takes a very long time to implement new items on existing surveys (and even longer to mount new surveys or survey components). Are there ways NCES can identify new issues earlier and collect information on them before building them into large-scale surveys? The FRSS is one option to collect data of this sort, but there may be others. What might serve as an “early warning system” for identifying potential upcoming information needs?

Summary

In summary, I see these papers as stimulating NCES to take a broader view of its role. NCES will not pursue all of the topics mentioned in the papers; that is not fiscally possible nor appropriate. However, NCES could think about who might provide such information to the American public (and how this might occur), and could play a role in seeing that it happens. The NCES role would vary by topic and activity, sometimes serving as a facilitator or coordinator. That would involve a new role for NCES, acting as a broker of information, identifying holes, and getting others to fill them rather than doing it directly (nor would NCES necessarily have to release the information, but rather monitor that someone does). It also implies working more closely with a wider group of actors, starting with colleagues in other parts of OERI and the Department of Education, but also reaching out to private groups, associations, foundations, business, and other interested parties.

Notes

1. Another is home schooling. In the October 1994 Current Population Survey (CPS), items were included on home schooling. Out of the nearly 60,000 households in CPS, there were about 100 children between the ages of 6–17 who were being home schooled.

2. A paper submitted to NCES subsequent to the Futures conference (Baker 1996) looks very much like this. In making recommendations about changes to SASS that would help monitor reform efforts and their impacts, the author suggests reorienting SASS so that its primary focus would be on gathering information on schools as organizations. He cautions against focusing on particular

“reforms,” because they come and go (and come again) relatively quickly. “[T]he key is to think of ways to capture information about reform without being tied to any one particular trend over a lengthy time” (p. 31).

References

- Baker, D. 1996. “Towards an Organizational Database on America's Schools: A Proposal for the Future of SASS, with Comments on School Reform, Governance, and Finance.” Commissioned paper for the U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, Forum on the Design of the 1998–99 Schools and Staffing Survey (SASS).
- Medrich, E.A., Brown, C.L., Henke, R.R., Ross, L., and McArthur, E. 1992. *Overview and Inventory of State Requirements for School Coursework and Attendance*. Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

APPENDIX A

Topics and Strategies Mentioned in Two Papers

Jennings and Stark	Cross and Stempel
Topics	
Effect of reform on achievement	Effect of reform on achievement
Standards-based reform—what states are doing (updated), standards and assessments, OTL	Value added by reform
Graduation requirements	Effect of academic standards on schools
Charter schools—how successful	Charter schools—evaluate performance
Reform networks	
School choice	
Private, for-profit companies run schools	
State takeovers	
School finance, state-by-state	
Effect of reduction in federal funding	
Effect of going to block grants	
School facilities	
Vocational education—tech-prep, school-to-work	
Home schooling—state-by-state, oversight procedures	Home schooling
Postsecondary reform initiatives	
Postsecondary student achievement	
Outcomes other than achievement	
International comparisons with states	
Regional, state, local data on reform	
Information on demographic, economic characteristics of school districts	
Report on specific reforms at all levels	Degree to which reforms are being implemented
	International benchmarking
	Motives and expectations of other countries
	Coordination of curriculum and expectations within schools and across schools in district
	Student mobility
	Interaction of elements of reform
	Technology—use, infrastructure
	Magnet schools
	Vouchers

Site-based management—many different kinds; impact on learning
 What occurs in classroom
 Quality/uniformity of assessments/testing in classroom; data on teacher grading schemes
 Cost/benefits of various reforms
 How successful schools streamline use of limited resources

Strategies

Broadest definition of reform to remain impartial
 Advisory committee—what reforms; priorities
 Annual report—“encyclopedia of education reform,” state activities, include data of others
 Other sources of financing—i.e., business and foundations
 Compile information collected by others; “repository” of others' data
 Other data on achievement—state assessments, college entrance exams
 Use state data for sub-state data
 Collect and report by state
 Put information up on Internet—includes research findings and updating data
 Make data widely available to researchers and others
 Determine what is education reform; common definitions
 Timely information, relevant, and easily accessible
 Concise reports suitable for policymakers
 Have information available for reauthorization; time publications to that schedule
 Most recent data on Internet
 More analysis of data—more interested in reasons than mere facts; why reforms work; need data over time
 Encourage analysis of reform data by researchers

More analysis of data

NCES/OERI conference on how to make data
more accessible and promote analysis
Contract out analysis

Integrate qualitative and quantitative
information about reform
Modify teacher part of NAEP to get
information about teacher education and
development (SASS)
Longitudinal information on teacher
careers
Collect data from multiple sources;
unreliability of self-reports
Independent data collection agents or
“inspectors” because self-reports are
unreliable
Link information at all levels to see what
is happening in specific classrooms
Extend purpose of data, more reliable,
more “subtly realized”
NAEP—expanded items to examine
schools' role in achievement
Data collections more delicate, sensitive,
focused

3

Curriculum, Pedagogy, and Professional Development

THIS PAGE INTENTIONALLY LEFT BLANK

Enhancing Opportunity to Learn Measures in NCES Data

**Dominic J. Brewer
Cathleen Stasz**

INTRODUCTION

What takes place in American K–12 classrooms? What is being taught, how is it being taught, by whom, and with what resources? Knowing the answers to these questions would seem to be a necessary information base upon which to build public policy aimed at boosting student performance, ensuring an equitable delivery of schooling for all students, and guaranteeing accountability of teachers and schools. The nation's school system has been the focus of much public discontent over the past decade, centered on perceived declines in student academic achievement, school inefficiency, and lack of accountability. Consequently, schools have been subjected to an unprecedented era of “reform,” ranging from changes in assessment to new curricula and graduation requirements to new models of school organization. Most of these changes are ultimately designed to bring about improvements in student outcomes, typically measured (narrowly) by standardized tests, via changes in what takes place within classrooms around the nation. However, given limited understanding of the determinants of student performance, the difficulties of measuring the inputs, processes, and outputs of schooling, and the many and disparate activities and clientele of the school system, systematically assessing the real impact of these reforms is no easy task. A precondition for this, however, is an accurate, detailed picture of what takes place in American classrooms.

A vast volume of research within education and other disciplines has attempted to map out and explain the processes of teaching and student learning. Within the classroom setting, the focus has been on curriculum content, pedagogical strategies and instructional goals, teacher characteristics, and other instructional resources. Recently, national data on such issues, based primarily on survey responses of teachers, have been collected by the National Center for Education Statistics (NCES). Typically part of large-scale national (often longitudinal) studies designed to meet a variety of diverse needs (e.g., researchers from many disciplinary backgrounds and policymakers interested in a host of issues), these data have a broad rather than a deep focus. They do, however, have the advantage of drawing on large sample sizes and are carefully designed and implemented. While this effort to collect data on curriculum and pedagogy at the national level is in its infancy, there is considerable doubt as to whether the complex nature of teacher and student behaviors, and their interaction in a classroom setting, can be captured by survey data.

This paper reviews attempts to date to collect classroom-level data, and discusses whether the mapping of the intricate, multidimensional activities of the classroom can be improved via better survey designs and instruments, or via other forms of data. The next section defines in more detail what we mean by curriculum and pedagogy, utilizing the concept of “opportunity to learn” (OTL). The following section discusses the rationale and uses for such data, and briefly outlines possible future needs in this area. The paper then presents an overview of existing NCES data collection efforts via national surveys. Several non-NCES major data collection efforts have been undertaken in recent years geared toward improving measures of OTL through a variety of alternative methods. These include the use of teacher daily logs, collection of classroom written assignments, test and texts, teacher interviews, classroom observation, and videotaping of classes. While this work is relatively new, it provides some potential avenues for future NCES data gathering. The paper concludes with a set of recommendations with regard to future efforts.

DEFINING “CURRICULUM AND PEDAGOGY”

We take an expansive view of “curriculum and pedagogy,” focusing on what takes place inside classrooms at the K–12 level in the broadest sense. This includes what is being taught (e.g., curriculum content); how it is being taught (i.e., pedagogical strategies); who is teaching (e.g., teacher—and sometimes student—characteristics); the instructional resources being used directly (e.g., textbooks); and the resources for teachers that support instructional goals (e.g., planning time, staff development, and opportunities for faculty collaboration). The activities of any classroom thus include not only the behaviors of teachers but also the activities and interactions among students, and between teacher and students. This whole gamut of classroom intentions, behaviors, and activities is obviously an extremely complex one with many dimensions. Therefore, obtaining a coherent and usable description of this picture is no easy task.

Opportunity to Learn (OTL)

One widely used way to organize thinking about curriculum and pedagogy is the concept of “opportunity to learn” (OTL).¹ Typically OTL research has divided classroom attributes into three distinct categories: curriculum content, instructional strategies, and instructional resources. Briefly, curriculum² refers to “the knowledge and skills the teacher presents for the students to learn during their classroom experiences” (NCES 1995b). This typically includes major and minor topic coverage, time spent, and more subtle teacher emphasis on topics. Instructional strategies refer to the ways in which teachers convey material to and engage their students. Traditionally, the latter has included issues such as the manner in which material is presented to students (methods, pace); questioning strategies; communication with students; expectations for students; classroom organization; grading and homework policies; allocation of time within a class period; and content organization. Instructional resources include basic learning materials such as books and supplies; equipment (e.g., computers); and the physical classroom environment (such as heat, light, furniture, and so on). We also include teachers' knowledge and preparation of curriculum and strategies under this heading (sometimes it is classified under curriculum). For more complete descriptions of the elements of OTL, see NCES (1995b).

While the OTL framework is a useful one, we believe it is important not to view the three aspects of OTL as rigid. Indeed, developing a richer and fuller picture of classroom activities hinges on being able to successfully identify and measure the *interactions* and overlap between curriculum, pedagogy, and resources, and their effects on learning. It should be clear that many classroom activities defy simple categorization; for example, as Porter (1991) has pointed out, separating curriculum content from pedagogy is tricky: “If a teacher uses story problems to teach problem solving, but all of the story problems involve the same format and the same operation for solution, then after the first few problems, the task becomes one of drill-and-practice for skill, not one of application and problem solving” (p. 18). Similarly, Burstein et al. (1995) note the interaction between content and teaching strategy and inability of survey data to capture “subtle differences in how teachers define and used different techniques” (p. 36). Thus, while one teacher might draw heavily from the textbook and do most of the talking in class, and another might use other sources and engage students in lively exchanges, both are likely to report these activities as lecturing.

Student Learning, Teacher Doing

In setting our focus on classroom activities in broad terms, we deliberately wish to call attention to what we believe is a neglected part of many analyses of curriculum and pedagogy: *the process of student learning*. NCES and most other data collection activities emphasize assessment of *teacher* behavior whether that stems from traditional process-product research or from newer, “reform-oriented” pedagogical approaches, such as the NCTM standards (Burstein et al. 1995). Clearly, however, teachers represent only one part of the classroom. This emphasis on teaching is consistent with instructional design theory (cf. Reigelut 1987). From this perspective, the task is to design an instructional delivery system that transmits content and skills in a clear, well-structured, and efficient manner. The approach stems from behaviorist theories of learning, but also has assimilated aspects of cognitive research in recent years (Collins in press).

In contrast to an instructional delivery view, a constructivist view argues that education should help students construct their own understandings. This perspective leads to an emphasis on learning rather than teaching, and on facilitative environments rather than instructional goals. It implies an approach to education that looks very different from traditional instructional design theory, and where considerations of curriculum content, for example, do not hold center stage. Collins' discussion (in press) of design trade-offs among learning goals, for example, addresses goals for what students should learn: memorization versus thoughtfulness, whole task versus component skills, breadth versus depth of knowledge, and so on. Such considerations about student learning goals apply whether the topic is math, social studies, or electronics.

Recent NCES surveys (e.g., The National Educational Longitudinal Study of 1988, [NELS:88]), as well as non-NCES efforts (Burstein et al. 1995), include teacher survey items based on reform efforts that in turn reflect constructivist views of education (e.g., National Council of Teachers of Mathematics [NCTM] standards, and California curriculum frameworks). While these items (discussed in more detail later) are meant to assess teachers' use of newer practices, they still focus on what *teachers* do.

While recent OTL-oriented studies have recognized and attempted to rectify the failure to assess student learning (Smithson et al. 1995), these efforts are still too new to determine if they validly represent student learning processes. Studies of how students experience the curriculum or how students think about what teachers do seem to be one starting point for examining the student learning process. Unfortunately, we currently know little about the varieties of student experience in classrooms, and what we do know has not been integrated into recent theoretical orientations toward teaching and learning (Good 1995; Erickson and Shultz 1992).

WHY COLLECT NATIONAL DATA ON OPPORTUNITY TO LEARN?

Why collect national data on opportunity to learn? We believe obtaining basic information about the nation's classrooms is a useful, necessary, and important undertaking for both policymakers and the research community. The extent to which we are able to describe the current educational system is surely a crucial element in efforts to improve it (Stecher 1992). Several reasons for this position are discussed in this section of the paper. For example, fundamental questions about effective practice—what teacher behaviors promote student learning—remain unanswered. Policymakers interested in reaching informed decisions about school reform need to know how opportunity to learn is distributed across students and with what effect, and how much change at the classroom level really takes place as a result of ongoing reforms in curriculum, assessment, and teaching practice. Further moves toward increased teacher and school accountability may require more numerous and more refined indicators of classroom activity. Without a clear view of *why* data on curriculum and pedagogy are needed, and how these needs may change, future data collection is unlikely to generate useful insights and may divert needed resources from elsewhere. As Porter notes, “the value of an indicator of school processes is determined by its problem orientation and policy relevance” (1991, p. 23).

What Promotes Student Learning?³

Prompted by unfavorable international comparisons, the overarching interest of policymakers over the past decade has been to improve American students' educational attainment, typically as measured by standardized test scores. Following a lengthy period in educational research and practice in which “inputs” to the schooling process and how these inputs were distributed across types of students were the focus of attention, emphasis switched in the 1980s to a strongly outcome-oriented paradigm. Raising educational productivity is clearly an important goal for the nation's schools. To the extent that researchers could identify the ingredients of successful classrooms—attributes of teachers or curriculum or classroom resources, for example—which are generalizable across different settings, a formula for improving student outcomes would be found. This “recipe” could then be applied throughout the nation in order to improve outcomes.

In fact, this line of research, either in the “process/product”⁴ or “effective schools”⁵ genres within educational research, or in the “educational productivity”⁶ literature within economics, has been unsuccessful in arriving at strong conclusions in regard to “what works.” Process-product research has revealed, for example, correlations between the pace of instruction, how information is presented and teacher's questioning strategies, and student outcomes. Work on curriculum content

has generally shown some link between the number of courses taken in a particular subject and achievement in that subject. The effective schools' literature suggested that "instructional leadership," clear school goals, and high expectations were particularly important in promoting student outcomes. The productivity paradigm has led to fewer robust conclusions, though there is a dispute over the interpretation of the numerous studies. In general while crude measures of teacher "ability" have been shown to be important determinants of student achievement, holding a set of school and student characteristics constant, other indicators such as teacher experience and degree levels have shown an inconsistent relationship to test scores.

The reasons for the rather disappointing results from this research are numerous. First, "outcomes" are much broader than standardized tests, which are often poorly designed and inadequate measures of student learning. Since outcomes such as problem-solving skills, self-esteem, communication skills, and citizenship are far harder to assess in a systematic and measurable fashion, they have tended to be ignored or relegated to footnotes by researchers. Pressure from the public to achieve tangible gains in measured performance has led to test scores being used as "political symbols as much as assessment devices" (Good 1995, p. 4).

Second, much of the process-product research has been based on small unrepresentative samples and has focused on one particular factor or part of classroom activity; in general, it has utilized univariate statistical methods in which causality is impossible to infer with confidence. Available national data for use in more sophisticated statistical studies, while improving, remain crude. Only recently, for example, have student tests on national surveys been explicitly tied to curriculum content and their teachers and classrooms. Most of this research has been conducted by economists who have focused almost exclusively on resource inputs (such as classroom size, or expenditures per student), and has ignored the more subtle and intangible multitude of teacher behaviors (such as whether a particular teacher lectures, or teaches a certain number of units of algebra). There have been few attempts to combine the two strains of research to produce a firmer basis upon which to draw conclusions as to "what works" in terms of classroom activities (see Murnane and Phillips 1981).⁷

Third, and perhaps most fundamentally, it is a debatable point as to whether such outcome-oriented analyses can provide a recipe that educational policymakers could adopt given differences across schools and classrooms. While it is possible, for example, to identify student and teacher classroom characteristics and crude indicators of curriculum and instructional strategies, the ways in which these elements interact in any given classroom on any given day are likely to be very complex. Indeed, some researchers have argued that the multidimensional nature of both inputs and outputs in the classroom makes this line of research impotent. If idiosyncrasies dominate, a formula that could be implemented across classrooms and schools will never be found (Monk 1992).

While these problems associated with answering the question "what promotes student learning?" have led to few strong conclusions, it is probably premature and overly pessimistic to abandon this line of research altogether. Indeed, more recent work focused on the impact of curriculum content on achievement has been more successful (see the brief overview in Burstein et al. 1995, pp. 3–5). Increased attention on data needs in the curriculum and pedagogy area is partly a result of earlier research failures and a desire to get inside the "black box" of schooling in a systematic way with nationally representative data. There seems to us to be further opportunities to

combine modern statistical techniques with improved and more refined national data on classroom activities and resources. Although it seems unlikely that a “silver bullet” will be discovered that can be used to cure the nation's educational productivity problem, even if better measures were to be collected by NCES, there is some merit in attempting to further refine measures of classroom activities in order to gain a better understanding of the ways in which curriculum content and instructional strategies may be related to student outcomes.

Do All Students Have an Opportunity to Learn?

Although somewhat de-emphasized in recent policy debates, descriptions of what takes place in the nation's classrooms are useful from an equity perspective. Ensuring educational equity is an important goal for the nation's schools. One particular emphasis in recent years has been on developing “indicator systems,” a key component of which is information on school processes (see Shavelson et al. 1987; Stecher 1992; McDonnell 1995). The main purpose of these efforts is primarily to provide data to policymakers (Porter 1991). OTL provides a way to determine if different students have equal educational opportunities. Equal opportunity is a fundamental concept in public schooling in the United States. To the extent that a certain allocation of educational opportunities across students is desirable, and that educators should strive to achieve such an alignment, it is necessary to first provide a description of the distribution of OTL. The growing interest over the past decade in obtaining indicators of classroom processes stems partly from a recognition that crude input measures are inadequate for assessing educational opportunity. Traditionally, resource (dollars per pupil) differences across school districts (Kozol 1991) have been emphasized, given the perceived link between spending and student outcomes. As is well known, this has led to numerous attempts in many states to equalize spending where inequities exist led by the California *Serrano* court decision in 1971 (*Serrano v. Priest* 1971). Equalization, such as it has occurred, has not led to an equalization of educational opportunity.

OTL offers a much richer process-oriented description of the schooling students receive. For example, the concept is based, in part, on a link between students' curricula exposure and their achievement (McDonnell 1995). Clearly, if students are not taught particular mathematics topics, for example, they cannot hope to score well on tests of those topics. NCES recently examined the link between course taking and achievement in math and science for students from different social backgrounds (Hoffer et al. 1995). Hoffer's analysis found that students from higher socioeconomic status (SES) families complete more courses in these subjects. In addition, students who complete more math and science courses show greater achievement score gains during high school, regardless of gender, race–ethnicity, and SES. Thus, additional coursework pays off for all students (see also Jones et al. 1986). The demonstration of a link between curricula exposure and outcomes has not been lost on policymakers; for example, Goals 2000 gives inducements to states to establish curriculum and student performance standards (Burstein et al. 1995, p. 5). Mapping course-taking patterns, then, is an important way to assess opportunity to learn. Further, there is already a good deal of evidence demonstrating considerable inequity in the distribution of classroom opportunities. For example, low SES and minority children tend to be taught by less qualified teachers, and have less access to tangible instructional resources such as computers (see Guitton and Oakes 1995, p. 324, for citations).

Distribution of opportunity to learn remains an important issue for public policy. It is not clear whether we know, for example, if “opportunities to learn are significantly different for students in Seattle as compared to those in Indianapolis” (McDonnell 1995, p. 310). On a more subtle level, information about the distribution of curricula offerings or instructional practices has led to efforts to reorganize course taking and classroom organization, for example, in “de-tracking” reform efforts. Several studies have shown that curriculum decision making in schools can match students with curriculum in ways that limit course taking and have implications for what students are exposed to and learn, and in some cases, their future educational opportunities (Oakes et al. 1992). In many states, increasing immigration raises concerns about the school's ability to cope with the educational needs of immigrant students and to keep them in school. OTL-type measures of what takes place within classrooms are likely to be informative about developing educational policies that can most effectively help such children. While use of OTL indicators for assessing equity is not without problems (as Guitton and Oakes [1995] argue, for example, different conceptions of equity lead to rather different emphases in developing OTL measures), it does provide an additional rationale for collecting data on curriculum, pedagogy, and instructional resources.

Are Current School Reforms Being Implemented?

A goal of many ongoing reforms being implemented in America's schools is to change curriculum and pedagogy in the nation's classrooms. It is difficult, however, to know whether reform rhetoric is translated into measurable change at the micro level: how change is implemented, the speed of implementation, and the barriers to change. Implementation of these reforms is a highly localized endeavor. Therefore, one important purpose of national data collection efforts on curriculum and pedagogy that is particularly important in an era of unprecedented change in schools is as a mechanism for monitoring that change at the level where it is being implemented.

In principle, descriptive information can provide a means of mapping out the type of changes that are taking place in schools, and where these are occurring. Further, with sufficient detail and coverage, survey data can be used to determine if indeed reforms are translated into positive discernable outcomes via non-experimental analyses. The National Assessment of Vocational Education (NAVE) provides one example where survey data have been used to track national reforms in vocational education. The most recent NAVE examines the extent to which states and localities are implementing specific program improvements mandated by Congress in the 1990 Amendments to the Carl Perkins Act (NAVE 1994).⁸

However, tracking reform via national surveys is no easy task given that change takes time and implementation is often slow. Surveys can only provide a snapshot of what is taking place at a point in time. Hence, it may be difficult to monitor change during a transition period; simply finding the appropriate words for survey items that adequately describe what should be taking place is problematic (Burststein et al. 1995). Given the pressure for certain reform it is impossible to know without additional information whether respondents simply adopt the current popular view about what they *should* be doing. This may be the case, for example, with the current national standards movement. Some of these issues are considered elsewhere.

Inherently tied to current reforms and the outcome-based emphasis of the past decade are increasing calls for accountability of teachers and their schools. Blank (1993) stresses that the focus of education reformers on accountability has increased interest in indicators of school processes (not the least via committees sponsored by the National Science Foundation, National Research Council, and NCES, among others). Hence, one additional rationale for collecting data on a national level on OTL is what McDonnell (1995) calls “high stakes” uses. OTL standards might be used, for example, in assessing whether or not schools meet certain practice standards and hence may be subject to sanction, i.e., using “force action in prescriptive ways” (Guitton and Oakes, 1995, p. 325). This view has also been expressed by O’Day and Smith (1993). At the very least they might be used in conjunction with other data in any system of school rewards based on outcomes (student performance). Clearly, any measures used for such purposes require a high degree of reliability and widespread acceptance from educators as legitimate measures. Hence, the use of OTL-type measures for accountability purposes has generated considerable controversy (Porter 1995; McDonnell 1995; and Guitton and Oakes 1995). NCES data are unlikely to be used for such purposes.

EXISTING NCES DATA: SCOPE, METHODS, AND LIMITATIONS

NCES and its predecessors have collected limited data on curriculum and pedagogical practices in American schools for some time.⁹ The primary mode of data collection has been to include items relating to classroom activities on large national surveys completed by principals or other school administrators and by teachers. These items have been primarily limited to information about teachers’ educational backgrounds, and school or classroom resource indicators (such as number of pupils per teacher). Some recent major NCES surveys include High School and Beyond (HSB), The National Educational Longitudinal Study of 1988 (NELS: 88), the Schools and Staffing Survey (SASS), and the ongoing National Assessment of Educational Progress (NAEP).¹⁰ Detailed information on classroom processes including curriculum content and instructional strategies is relatively new to NCES surveys; NELS was the first major survey to contain a considerable set of questions relating to specific curriculum topics within major subject areas and items focused on teaching behaviors within a particular classroom. Stecher, in reviewing available measures of curriculum content in 1992, concluded that very little was known about topic coverage or instructional methods (1992, p. 56). Further, only NELS has systematically combined longitudinal classroom-level survey data on students and their teachers that include student test data.¹¹ Hence, national data collection in this area must be considered to be in its infancy.

Typical NCES teacher and school administrator survey coverage is summarized in Table 1. This table is not intended to be exhaustive but provides an illustrative introduction to data on OTL. We have categorized items into three groups based on an OTL framework: curriculum content, instructional practice, and instructional resources. The four major surveys noted above are included in the table.¹² In general, the table illustrates that only NAEP and NELS contains classroom-specific data on detailed curriculum topic coverage and teachers’ instructional strategies. Even instructional resource measures have been limited and confined to the school rather than classroom level, obscuring variation between classrooms within a school.

Table 1—Items on opportunity to learn in selected NCES national teacher/school surveys

	Curriculum content	Instructional practice	Instructional resources
I. National Assessment of Educational Progress (NAEP) Annual 1969-1980; then biennial; student and teacher components	Broad topic emphasis within subject area planned for academic year for each class; type of skills taught	Broad instructional methods, type and frequency of student tasks assigned during class; amount of homework; type and frequency of assessment; use of resources	Class size; access to resources (e.g., calculators); teacher education and training, including in subject-specific content areas and teaching techniques
II. High School and Beyond (HS&B) Longitudinal student surveys 1980 (10th and 12th graders), 1982, 1984, 1986, and 1992 (sophomores only); school survey (1980, 1982); administrator/teacher survey 1984	None	General teaching goals, class time allocation, homework and assessment strategies, student recognition	Average class size and ability, school level expenditures/pupil, teacher credentials
III. Schools and Staffing Survey (SASS) Teacher, administrator, and school surveys 1987–88, 1990–91, 1993–94	Broad topic areas within subject for each class taught	None, general attitudes/perceptions toward teaching	Class size, ability level for each class taught, otherwise none
IV. National Educational Longitudinal Study of 1988 (NELS: 88) Longitudinal student surveys 1988 (8th grade), 1990, 1992; teacher surveys (tied to students) 1988, 1990, 1992; school surveys 1988, 1990, 1992; parent surveys 1988, 1992	Topic emphasis within subject area; general concepts specific to subject area	Type and frequency of instructional method used; allocation of time to whole group	Class size; teacher credentials, including coursework in subject areas

SOURCE: Compiled by authors from respective survey instruments. See text for discussion.

HS&B was a national longitudinal study designed as a followup to the National Longitudinal Study of the High School Class of 1972 (NLS-72). The basic HS&B data consist of two cohorts (sophomores and seniors) of high school students initially surveyed in 1980. Basic school information was provided by principals, including some information on general school policies and instructional resources (class size, dollars per pupil, teacher credentials, and so on). A supplementary Administrator and Teacher Survey was conducted in 1984 that included separate surveys for principals, guidance counselors, and teachers. While teachers were not linked to particular students or asked about a particular class, they were asked a number of questions regarding the extent to which they controlled their classroom resources, curriculum content, and teaching techniques; how often classes were interrupted “on an average day”; the importance of general goals (such as “academic excellence,” “good work habits,” and “discipline,” and “moral or religious values”) in their teaching; allocation of classroom time between “daily routines,” “getting students to behave,” and “instruction or student practice of skills”; and detailed items relating to their qualifications and background. Clearly these items are unspecific and general in nature.

SASS consists of a set of surveys conducted in 1987–88, 1990–91, and 1993–94. The latest wave consisted of a teacher questionnaire sent to 65,000 teachers, along with administrator and school components. No student data were collected. The teachers were asked about classes taught during the most recent full week at the time the survey was completed. Broad curriculum topics—for example within mathematics 11 topic area codes such as “general mathematics,” “trigonometry,” “calculus”—class size, grade level, and ability level were included, but no specific items on instructional strategy. Teachers were asked about their backgrounds for teaching specific subjects.

NAEP’s “The Nation’s Report Card” was started in 1969 as a way of tracking the educational performance of the nation’s school children. Students in different grades are tested in a variety of subjects and their background information is collected. In addition, their teachers are surveyed to gather data on their background and instructional practices. NAEP continues to be conducted every 2 years, with samples of 4th, 8th, and 12th graders being tested since 1988; math and reading are assessed every 2 years, science and writing every 4 years, and other subjects less frequently (NCES 1995a). Curriculum content is surveyed with broad indicators for each subject. For example, 8th-grade mathematics teachers were asked in the 1992 NAEP: what emphasis they *plan* to give (“heavy,” “moderate,” or “little”) to five topic areas during the course of the academic year: “numbers and operations,” “measurement,” “geometry,” “data analysis, statistics, and probability,” and “algebra and functions.” Instructional strategies are also assessed via one or two items. For example, 8th-grade writing teachers were asked in the 1992 NAEP: “Do you use any of the following instructional approaches?” “Grammar or skill-based instruction,” “writing process instruction,” “integrated reading and writing,” “writing about literature,” “writing across other subject areas.” A threefold scale (“Yes, as a central part of instruction,” “Yes, as a supplement to instruction,” and “No”) was utilized.

NAEP also assesses student activities. On the 1992 NAEP, for example, 8th-grade math teachers were asked: “How often do the students in this class do each of the following things?” Eleven activities were assessed—such as “do a mathematics problem from their textbooks,” “do a mathematics problem on worksheets,” “solve mathematics problems in small groups,” “discuss solutions to mathematics problems with other students”—on a scale of “almost every day,” “once or twice a week,” “once or twice a month,” or “never or hardly ever.” Interestingly, questions relating to these issues are not confined to teachers; for example, *students* are asked: “How often does the

teacher . . . read aloud . . . do a problem on the board?" In the past, however, NAEP has sampled students and teachers in such a way that only a subsample of a teacher's students were sampled.

NELS is perhaps the most detailed and most recent collection effort. A group of 1988 8th graders have been surveyed and tested in 1988, 1990, and 1992, along with detailed survey information from these students and their teachers in at least one subject-specific class. Additional school-level and parent information was also collected. NELS is thus unique in that it permits researchers for the first time to link a national sample of students with the teachers who actually taught them. Further, the standardized tests administered to the students were, in contrast to HS&B, for example, linked to the curriculum studied by the students. The teacher surveys contain detailed information on the curriculum content used in a particular class, instructional methods used, and the teacher's goals for the class, in addition to the standard range of items on teacher educational background, class size, and school-level resource measures. Some examples of specific items taken from the 1992 mathematics teacher survey are given in Table 2 in order to provide an indication of the level of detail of these survey items.

The curriculum content items include both specific topic coverage and degree of emphasis on different objectives for mathematical learning. The list reflects newer conceptions of appropriate instructional goals for teaching mathematics (e.g., learning to represent problem structures in multiple ways). Similarly, the instructional practice items attempt to discern the percentage of time engaged in teaching (individuals, groups, whole class, labs) relative to other non-instructional activities, such as maintaining order. Items assessing teaching methods attempt to distinguish between teacher-centered activities (e.g., lecture, lead group discussion), and more student-centered activities (e.g., give oral reports, work in cooperative groups), and the frequency of these activities and of uses of instructional media. Like the content and emphasis questions, these methods questions appear to reflect "reform" practices espoused by mathematics educators in such documents as the NCTM Professional Standards (1991).

National survey data collected by NCES have several major benefits: sample sizes are typically large, nationally representative, and carefully designed and implemented. Hence, these data have a degree of generalizability that other research efforts cannot match. The data collected by NCES contribute in large part to the goals of data collection in this area outlined earlier. For example, NELS affords researchers the opportunity to explore the relationships between curriculum content, instructional strategies and resources, and student achievement in a particular subject area, while controlling for school and other contextual factors and students' prior ability level. Hence it is a remarkably rich data source. NAEP provides a reasonably comprehensive snapshot of the curriculum in math and reading across the nation. However, NCES data collection efforts are geared toward the many diverse groups that the agency is charged with serving, resulting in a broad-brush approach, rather than one focused more deeply on specific areas such as curriculum content at the K-12 classroom level.¹³ Hence curriculum content indicators, while becoming more detailed, remain at a relatively general level with a handful of topic areas identified within a subject. Instructional behavior is captured in crude terms both in terms of strategies identified and the response scales used. Surveys have major limitations for collecting very rich information on curriculum and pedagogy.

Table 2—Examples of teacher survey items from NELS Second Follow-up

Curriculum Content

Have you taught or reviewed the following topics in this math class during this year?

Integers; patterns and functions; linear equations; polynomials; properties of generic figures; coordinate geometry proofs; trigonometry; statistics; probability; calculus.

Scale: No, but it was taught previously; Yes, but I reviewed it only; Yes, I taught it as new content; No, but I will teach or review it late this school year; No, topic is beyond the scope of this course.

How much emphasis do you give to each of the following objectives?

Understanding the nature of proofs; memorizing facts, rules, and steps; learning to represent problem structures in multiple ways; integrating different branches of mathematics; conceiving and analyzing the effectiveness of multiple approaches to problem solving.

Scale: None, Minor, Moderate, Major

Instructional Practice

What percent of your time did you spend?

Instructing whole class; instructing small group; instructing individuals; maintaining order; administering tests; administrative tasks; conducting lab periods.

Scale: none; <10%, 10–24%, 25–49%, 50–74%, 75–100%

How often do you use the following teaching methods or media?

Lecture; use computers; use audiovisual material; have teacher-led whole group discussion; have students respond orally to questions on subject matter; have student-led whole group discussion; have students work together in cooperative groups; have students complete individual written assignments or worksheets in class; have students give oral reports.

Scale: Never, 1–2 times a month, 1–2 times a week, almost every day, every day

SOURCE: NELS Second Follow-up Teacher Survey (1992).

Although non-NCES studies have begun to collect data in a wide range of different ways (as discussed in the next section), NCES has to date relied solely on survey instruments (and largely on *teacher* surveys) as a means of collecting information on curriculum, pedagogy, and instructional resources. There are several reasons why this survey approach may be inadequate for collecting information on these areas. First, given that a respondent's time is not costless, the number of items that can be devoted to these topics is necessarily limited in multi-purpose surveys. Given the complex and multi-faceted nature of classroom activities, though, it is not clear whether sufficient useful information can be gathered in a few items, particularly on instructional practice and goals.

Second, there are serious validity issues arising in the use of general survey items. Since classroom activities vary from class to class depending on the subject, groups of students present on any one day, and student and teacher moods vary, “each section of each course results in a potentially unique content description” (Porter 1991, p. 15). Surveys are inherently static (i.e., conducted at a point in time), and continual traditional written surveying is evidently costly and impractical. But since classroom activities vary from day to day and group to group, key issues arise as to the timing of survey instruments and the reference point for teacher responses. Even if there were no variation problem, interpretation of specific activities and events is subjective. The same classroom activities from the perspectives of student, teacher, and outside observer may be very different, and NCES studies primarily rely on teacher perceptions alone.

Third, although recent broad-based national surveys such as NELS have greatly improved classroom-level information, the items tend to be descriptions of what teachers do, rather than what students do. While we may know from the handful of questions teachers are asked about curriculum content and instructional strategies that, for example, a particular classroom in 8th-grade math is lecture-based, has a certain number of units of algebra, and has two computers, these three elements could be combined in numerous ways to produce differing classroom environments and learning opportunities. Similarly, existing data tend to be based on a particular conception of the teaching process, one based on a direct-teaching, whole-classroom model. This may be inappropriate in a dynamic educational world in which many reform efforts are seeking to change important aspects of curriculum and classroom organization.

Finally, NCES surveys have collected only limited information on instructional resources, emphasizing class size, student ability, and teacher credentials. As Oakes and others have shown, school context factors can significantly affect classroom instruction and students’ opportunity to learn (Guitton and Oakes 1995; Stasz et al. 1990, 1993). An important factor for interpreting survey data on the implemented curriculum and teacher practices is professional teaching conditions. If teachers receive limited staff development to enhance their repertoire of teaching techniques or learn about new approaches for teaching mathematics or other subjects, then it may not be surprising to learn that they rely on “traditional” methods in their teaching. Put another way, lack of instructional resources, including materials (e.g., computers, textbooks) and professional development, may inhibit “opportunity to teach.”

IMPROVING DATA ON OPPORTUNITY TO LEARN

Up to this point, we have argued that the concept of opportunity to learn is a useful way to encompass curriculum content, pedagogy, and instructional resources in schooling. We also argue that measures of OTL can be enhanced by placing a greater emphasis on student learning, not only through tests that assess the “attained” curriculum but also by gathering information on the learning process. We further argue that while teacher survey items may provide a reasonable description of curriculum content, it is doubtful that they can ever satisfactorily assess detailed elements of classrooms without undue burden on respondents; hence other forms of data collection that are available might be considered by NCES. The goal is to be able to provide researchers and policymakers with a richer picture of what takes place within classrooms both in terms of how students learn and what students and teachers do.

In this section, we discuss various ways in which NCES data collection could be improved that draw on several recent efforts to assess OTL and that acknowledge the shift from a process-product orientation toward one that emphasizes teaching and learning for understanding. Our discussion is grouped into three sections: 1) enhancing measures of curriculum content and instructional practice items on national teacher surveys; 2) enhancing measurement of student learning processes through student surveys and other methods, such as observation, artifact collection, interviews, video data, teacher logs, and so on; and 3) enhancing instructional resource measures on surveys.

In discussing possible improvements, we draw on the contributions of several non-NCES efforts that in recent years have attempted to collect a broader array of data on opportunity to learn using a wide variety of alternate tools. These include the Reform Up Close project, conducted by the Consortium for Policy Research in Education; the Validating National Curriculum Indicators project, conducted by a team of RAND/University of California-Los Angeles researchers; and the Third International Mathematics and Science Study (TIMSS).¹⁴ The components and data collection methods used in these efforts are summarized in Table 3. The purposes of each study differ somewhat and the methodologies used reflect this. It should be stressed that these efforts are ongoing at least in the sense that analyses of the findings are preliminary; hence any conclusions drawn from this work must be regarded as tentative. It will be several years before the data collected from these projects are fully evaluated. It seems likely, however, that they will provide valuable insights into possible enhancements to NCES data used to assess opportunity to learn.

As the name suggests, TIMSS is a cross-country study (about 50 countries in total) designed to evaluate teaching and learning in mathematics and science for 9- to 13-year-olds. It builds upon earlier similar studies (the First and Second International Mathematics Study, FIMS, and SIMS). Pilot testing for TIMSS was conducted in 1993 and 1994, and collection and analysis of data are ongoing. TIMSS includes a student achievement component, an assessment of student attitudes and background, and class-level data on opportunity to learn from teachers and school-level officials. In addition to survey data, textbooks/materials are being collected.

The RAND/UCLA study sought to assess the validity of data collected in national studies such as NELS. Its purpose was to “design and pilot a model for collecting benchmark data on school coursework” (Burstein et al. 1995, p. 1). Conducted using a small sample (just 70) of the mathematics teachers surveyed in the 1992 NELS sample of 12th-grade math teachers, the projects’ preliminary findings have been published (Burstein et al. 1995). Enhanced versions of various NELS teacher survey items were administered to these teachers (see next section). The project relied heavily on the collection and coding of artifacts such as classroom assignments, quizzes and exams, and textbooks, as well as daily teacher logs and some teacher interviews (originally unplanned).

Finally, the Reform Up Close project, funded by the National Science Foundation, compiled data on secondary mathematics and science in high schools in 12 school districts, using an array of different collection strategies. Since the focus of the study was an investigation of state and local reforms in math and science, interviews were conducted with state- and-district level administrators, as well as school principals and teachers. A teacher survey was administered

Table 3—Selected recent non-NCES studies assessing opportunity to learn

	Curriculum content	Instructional practice	Instructional resources
Third International Math and Science Study (TIMSS)	Topic coverage (text, teacher survey); time on topics (teacher survey); emphasis on topics (teacher survey); topic test items (teacher survey)	Teaching practices (teacher survey); student activities (teacher survey); classroom management (teacher survey); grading and homework (teacher and student survey); planning time (teacher survey)	Teachers' knowledge of topics (teacher and student survey); texts, equipment, facilities (teacher survey)
Validating National Curriculum Indicators (RAND/UCLA) (Burstein et al. 1995)	Topic coverage (teacher logs, artifacts, text, teacher survey); time on topics (logs, artifacts, text, teacher survey); emphasis on topics (logs, artifacts, text, teacher survey); topic test items (artifacts, teacher survey)	Teaching practices (logs, artifacts, teacher survey); student activities (logs, artifacts); classroom management (logs); grading and homework (artifacts, teacher survey)	Teachers' knowledge of topics (teacher survey); texts, equipment (teacher survey)
Reform Up Close (CPRE/RUC)	Topic coverage (logs, texts, observation, interview, teacher survey); time on topics (logs, observation, teacher survey); emphasis on topics (logs, teacher survey); topic test items (logs)	Teaching practices (logs, observations, interviews, teacher survey); student activities (logs, observations, teacher survey); classroom management (observations); grading and homework (logs, teacher survey)	Teachers' knowledge of topics (interviews, teacher survey); texts, equipment (logs, observations, interviews, teacher survey)

SOURCE: Adapted by the authors from NCES (1995b), Table 1-Summary, pp. 35–38. See NCES (1995b) for full references used to compile the information in this table.

to around 400 teachers; a subset of these (between 62 and 82) were observed by the researchers; and daily activity logs were completed. “The effort to represent ‘opportunity to learn’ in the classroom was an important part of the larger study but not the primary focus” (NCES 1995b).

Enhancing Measures of Curriculum Content and Instructional Practice Items on National Teacher Surveys

One relatively straightforward extension of existing NCES data efforts would be a revision and extension of existing NELS-type items on curriculum content on teacher surveys. Some guide as to how this might be done is provided by the three studies cited above, in particular the RAND/UCLA study that built directly on the NELS items. Table 4 shows some typical extensions of the NELS items used by this study, based on the original teacher survey items reported in Table 2.

The curriculum content items probe for topic coverage at a greater level of specificity than previous surveys and also ask about the number of class periods teachers spend on each topic (although as few as 10–15 minutes counts as a “period” on their scale). The enhanced instructional practice items include strategies advocated in mathematics reform efforts. Burstein et al. (1995) also scaled and factor analyzed these items to see if they could meaningfully define instructional “repertoires”—instructional strategies that occur together. Such repertoires might provide a more coherent picture of instruction than simply reporting frequencies of teaching behaviors on an item-by-item basis. Although their analysis was hampered by a lack of variation in classroom practices across teachers in their sample, their approach looks promising for assessing instructional repertoires and might be used in future studies, particularly those that try to link repertoires to student outcome data.

The Burstein et al. (1995) validity study concludes that it is possible to add further, more refined topic areas to curriculum content items, and also additional questions on instructional strategies albeit with close attention to the response scale provided. The research team further recommends dropping items relating to instructional goals. Given the “paucity of empirical work regarding the definition and validation of curriculum-specific instructional constructs” (Stecher 1992, p. 76), the recommendation of Burstein et al. (1995) for future validation studies makes sense. They suggest that at the outset of large-scale national surveys, in-depth studies of small samples of teachers be conducted, using techniques that measure instructional processes with greater subtlety than is possible through survey data (Burstein et al., p. 56). These recommendations recognize that the language of instruction is in a state of flux, which may partly account for findings of lower validity for content and practices associated with the mathematics reform movement.

Earlier surveys relied on findings from process-product studies to identify “effective” practices to develop items (NCES 1995b), while more recent surveys have used state curriculum frameworks and reports from various professional groups, such as NCTM. Future studies to assess teaching practices should also look to current research on “teaching for understanding” (Good 1995); Blumenfeld’s research on science teaching (1992) or research on teaching from the

Table 4—Examples of enhanced NELS teacher survey items on curriculum and pedagogy from RAND/UCLA Study

Curriculum Content

Have you taught or reviewed the following topics during this year in class?

Estimation; proportional reasoning; tables and charts; graphing; math modeling; ratios; proportions and percents; conversion among fractions decimals and percents; laws of exponents; inequalities.

Scale: No, but it was taught previously; Yes, but I reviewed it only; Yes, I taught it as new content; No, but I will teach or review it late this school year; No, topic is beyond the scope of this course.

Indicate the appropriate number of periods devoted to each topic. If you focus on topic for 10 or 15 minutes on a given day, count that as a period.

Topic list above.

Scale: None; 1 or 2 periods; 3–5 periods; 6–10 periods; more than 2 weeks, but less than 1 month (11–20 class periods); 1 month or more.

Instructional Practice

How often do you use the following instructional strategies with this class?

Demonstrate working an exercise on the board; have student work on exercises on the board; use manipulatives to demonstrate a concept; have smaller groups work on problems to find a joint solution; have students work on problems for which there is no obvious solution; have students keep a mathematics journal; have students represent and analyze relationships using tables and graphs.

Scale: Almost every day; once or twice a week; once or twice a month; once or twice a semester; never.

SOURCE: Burstein et al. (1995)

constructivist perspective (e.g., Collins et al. 1989; Stasz et al. 1993) to identify effective teaching practice. Selecting items is not straightforward, however, as it is important to be able to capture the range and variety of traditional and reform teaching practices, rather than focusing on a narrowly defined vision of practice (Smithson et al. 1995).

Enhancing Measurement of Student Learning Processes Through Student Surveys and Other Methods

An obvious way to extend assessment of curriculum and pedagogy, and to include information about student learning, is to administer student surveys. One recent attempt to administer a student survey may be found in the Smithson et al. (1995) study of middle-level science. The purpose of the survey is to assess the enacted curriculum in science to help states participating in the project to assess OTL, interpret the results, and improve classroom practice. Students are asked about the frequency with which they experience 27 separate activities (see Table 5) and about their previous exposure to science. Teachers of these students are asked about their educational background, influences on the curriculum of the science class, how computers and calculators are used, and homework and grading policies.

Most interestingly, teachers are also asked about the instructional activities of the student that are directly aligned to the 27 student items: How often does an average student do these things in science class? Thus, the responses of teachers and students can be compared. Items for teachers and students emphasize three kinds of activities: acquiring information, using information, and extending information. Since questionnaires were field-tested in the 1994–95 school year, results from this study will be forthcoming. Their approach, however, looks promising for assessing student learning and its relationship to teaching practice and student outcomes.

Survey data have been the only type of data NCES has sought to collect on a national basis on opportunity to learn. Other researchers have utilized a wide variety of methods, and in recent years these have been tied to survey data on curriculum and pedagogy either as a means of determining the validity and reliability of survey items or as a research tool in their own right. These alternative forms include the following: collection of teacher lesson plans; written assignments, exams, and textbooks; teacher logs detailing classroom time allocation and tasks; observation and video of classrooms in action; and teacher interviews.

Burstein and his colleagues (Burstein et al. 1995), for example, gathered teacher assignments (homework, quizzes, classroom exercises, projects, examinations) as a way to validate teacher reports of their practices because they represent much of the curriculum presented to students. These items probe the types of performance teachers expect from students; for example, what percent of test items “require a critique or analysis of a suggested solution to a problem” or “require the application of concepts and principles to different or unfamiliar situations.” Similarly they ask how frequently teachers assign various types of homework, such as “gathering data, conducting experiments, working on projects” or “explaining newspaper/magazine articles.”

In keeping with the primary purpose of this study, the authors compare test or homework survey items to artifacts (actual tests and assignments) to assess the validity of survey responses. In both cases, agreement was low, suggesting that surveys are not very reliable for assessing what teachers expect of students, particularly for “more innovative” items that reflect a reform-oriented perspective (i.e., encourage student-centered activity and construction of knowledge). They also recognized, however, that the curriculum presented through these artifacts does not provide information about how students receive and respond to the curriculum. To get a sense of student

Table 5—Examples of student survey items on opportunity to learn

How often do you do the following activities in your science class?

Listen to your teacher or someone else explain things about science
 Read about science in books, magazines, or articles in class
 Collect data from sources in books, magazines, or articles in class
 Read tables, graphs, or charts
 Use measuring tool such as rulers, thermometers, balances, computers, and so on
 Do a laboratory activity, investigation, or experiment
 Observe experiments or investigations that others do, including teacher demonstrations
 Watch films or videos
 Use laboratory equipment
 Work in small groups
 Participate in school planned and supervised activities outside the classroom
 Work on assigned science projects or activities on your own away from school
 Use the computer in science
 Answer questions from your science book
 Take a quiz or test
 Write about science (e.g., lab reports, science papers)
 Make your own tables, graphs, and charts
 Change something in an experiment to see its effects
 Design experiments
 Ask questions to improve your understanding
 Make predictions, guesses, or hypotheses
 Make maps/drawings or models to show scientific ideas
 Reach conclusions about scientific data
 Choose a method for expressing an idea to your class
 Revise and improve your work
 Apply scientific concepts to your everyday life
 Explain what you learn in science relates to real-world issues (such as the environment)

Scale: Nearly every period; about once a week; once or twice a month; once or twice a year; never.

Teachers are asked about same list: *How often does an average student do these things in science class?*

SOURCE: Smithson et al. (1995). Items taken from middle-level science student survey.

learning, they asked teachers to provide examples of student work associated with each major assignment. This request appeared overly burdensome for teachers, however, and this data collection was subsequently abandoned (Burstein et al. 1995).

Teacher logs were collected by both RUC and RAND/UCLA studies as a way of mapping out the daily activities of teachers. The RAND/UCLA study was designed so that researchers were able to directly compare instructional practice survey items (such as those in Table 4) with teacher logs and artifacts such as homework assignments, quizzes and exams, and textbooks. The teacher logs were completed at the end of each day, collecting information on topic coverage, student activities, and modes of instruction. In keeping with the small-scale “benchmarking” purposes of the study, logs and artifacts were collected only during a 5-week period. The rate of agreement between surveys and logs was “quite low,” although part of the explanation for this finding “may lie in how the survey response categories were constructed” (Burstein et al. 1995, p. 39).

RUC used one-page daily logs in which teachers recorded lesson topic, subtopic, presentation mode, and student performance expectations using a 4-digit code; this permitted the coding of, for example, almost 6,000 different 4-digit content characterizations of math lessons (NCES 1995b). These data were collected from teachers over an entire year, and compared for some lessons with structured observer reports focusing on instructional activities, student engagement, and classroom management. While comparison of survey data with observations and logs showed only a moderate degree of agreement, observers’ reports and teacher daily logs for the same lessons “indicated that some dimensions of instruction could be described with a high degree of inter-rater reliability in an activity that takes only a few minutes a day” (NCES 1995b).

While surveys, artifact collection, teacher logs, and teacher interviews can provide a great deal of useful information about curriculum content and teacher activities that occur frequently and are well established, “some aspects of curriculum practice simply cannot be measured without actually going into the classroom and observing the interactions between a teacher and students” (McDonnell 1995, p. 310). This truth raises fundamental issues for national data collection efforts if richer data on classrooms are to be collected. Clearly there are limits to the extent to which outside observers can enter classrooms and assess lesson content, instructional strategies, and student activities, both in terms of cost and in terms of generating useful information. Observing lessons is labor intensive and hence very costly if done on a large scale. Using observations in validity studies, i.e., to check the response to survey data, may be confined to small samples, but even here there is a concern about the representative nature of the subsample of teachers observed. There are difficulties in deciding which classrooms to pick and when to observe them. This problem was encountered to a degree in the Burstein et al. (1995) study in which daily logs and artifacts were collected from just 70 teachers. It turned out that the background qualifications in mathematics of these teachers was considerably different from the wider NELS sample, raising doubts about the study’s overall findings.

Further, if data gathered through these non-survey methods are to be used for purposes beyond simply assessing the validity of survey items, they would ideally be generalizable to some extent. This implies the need for structured forms to record classroom observations and careful observer training so that similar behaviors are recorded as similar by different observers. RUC utilized a structured observers’ form as a means of assessing log and survey items, but also offered them an opportunity to provide a narrative report dealing with more subtle aspects of what was taking place in the classroom. It is not clear that methodological techniques exist to the point where such data could be coded or analyzed in ways that produce generalizable findings at reasonable cost.

One possible way in which the costs of observation and difficulties of interpretation may be reduced is through the collection of video data. The TIMSS Videotape Classroom Study currently under way is collecting information about classroom mathematics instruction to supplement data from assessments and questionnaires collected in the main TIMSS study (Stigler and Fernandez 1995; see also the paper by James Stigler for this NCES Futures Project). The project collected a random sample of approximately 100 TIMSS 8th-grade classrooms in the United States and Germany, and 50 in Japan. The tapes have been transcribed onto CD-ROM and linked in a multimedia database to translated transcripts of classroom speech in order to enable computer access to video data. Tapes will be coded to describe classroom instruction in the three countries and can be linked to survey data on classroom instructional methods. Stigler and Fernandez (1995) describe field-test study procedures and lessons learned thus far in collecting video data (including hiring and training videographers) and designing the multimedia database. The outcomes of this project will have important implications for judging the costs and feasibility of using video data to track and assess teaching practice, student process, and so on.

Enhancing Instructional Resource Measures Using Surveys

The non-NCES studies discussed here as well as others point to an inconsistency between the rhetoric of reform movements and the reality of teaching practice. For example, Burstein et al. (1995) found internal consistencies in teacher surveys on reporting instructional practices and goals. However, follow-up interviews with teachers revealed that teachers did not know what “math modeling” meant, even though it appears in the state math curriculum frameworks. The authors conclude that survey data may not be validly interpretable at a time when practice is in flux. Their findings also suggest that additional information about instructional resources—particularly teacher professional development—might improve our ability to interpret survey data. As discussed earlier, if teachers lack staff development opportunities or work in a school where the community of practice does not support *teacher* learning, then how are teachers to come to understand “transitional” curriculum content and instructional practices? At the very least, additional survey items might assess “opportunity to teach.”

Current surveys rely on teacher background characteristics as indicative of teacher quality and ask teachers to report degrees and credentials, undergraduate major, the subjects they teach, and the like. These indicators tie “quality” to knowledge of subject matter and credentials. In addition to this information, surveys might assess teachers’ knowledge of and opportunities to learn about innovations, with questions like the following: How many workshops or other professional development activities have you attended this year? How many focused on new curriculum and instructional practices in mathematics? How would you rate the usefulness of these activities . . . for improving understanding of mathematics teaching reforms? . . . for changing the way you teach mathematics? . . . for changing the kinds of assignments you give to students? How often are you able to discuss new ideas about mathematics teaching with other teachers at your school? Does your school and district support teachers who want to adopt innovative curriculum and teaching practices? (Also, see Smithson et al. 1995 for a teacher survey that includes questions about professional development opportunities.) Broadening teacher background measures in this way would help strengthen existing data on instructional resources.

CONCLUSIONS AND RECOMMENDATIONS

In this paper, we have used the concept of opportunity to learn to frame our discussion of national data gathering efforts on curriculum and pedagogy. We have discussed the purposes of data collection in this area, reviewed NCES data collection based on national surveys, and drawn on some very recent studies that have refined survey items and widely utilized non-survey modes of data collection. Given the preceding discussion, we conclude with a set of specific recommendations that NCES might adopt with regard to improving data on opportunity to learn.

First, as demonstrated by recent non-NCES studies, teacher surveys seem an effective and efficient way to gather information about course taking if the standard is knowing whether or not a topic has been taught, and if it has been taught over several periods or weeks. More reliable data can be obtained by asking teachers more specific questions about particular curricular topics. Adding more finely grained items to teacher surveys would thus appear to be sufficient for gathering national data on curriculum content. More instructional strategy questions could be added to teacher surveys; in particular, to expand the types of teaching practices to assess any shifts from traditional to “reform-oriented” pedagogy. (Research from the “teaching for understanding” paradigm provides a source for identifying new items.) However, while practice is in transition, it appears to be worth continuing validity studies. This way, future efforts can provide reliable estimates of changing teaching practice. Collection of non-survey data (such as teacher daily logs, teacher interviews, and observations), and use of exams, quizzes, assignments, and textbooks for small subsamples of survey respondents apparently provide useful additional information that help supply a richer picture of classroom practice and a means of assessing the validity of survey items. Methods of coding such data have been developed in recent years (for example, by the Burstein et al. team). The feasibility of using videotape data such as that collected by the TIMSS project should be closely monitored by NCES, and a similar effort should be considered in conjunction with a future national data collection such as the Early Childhood Longitudinal Study.

Second, enhanced efforts to assess student learning should be undertaken in future NCES work. Although newer studies have attempted to include items about teaching practice that reflect the constructivist view, they remain teacher-oriented. Adding a stronger student component might be accomplished by adding class-specific items to student surveys that mirror those on teacher surveys, perhaps with classroom observations to assess item validity. Further, the enriched data collected by non-survey methods, such as collection of artifacts and video data, may be utilized in aiding understanding of student learning. We recommend that NCES begin to explore the usefulness of these methods for collecting information of student learning processes as a supplement to survey data.

Third, further items on teachers’ background and preparation for teaching—opportunity to teach—could also reinforce existing teacher surveys. Since many curricula and pedagogical reforms place new demands on teachers, it is important to determine whether or not teachers are equipped to adopt and effectively utilize these new methods. Currently, measures on instructional resources at the classroom level are limited in their number and scope.

Despite its flaws, NCES has been a valuable asset to policymakers and researchers in helping to understand opportunity to learn and outcomes of schooling. Recent efforts in Congress propose

to do away with the Department of Education and give more control of federal dollars to the states. The shift to block grants to states can only enhance the diversity we already see in our nation's schools, per-pupil expenditures, teacher student ratios, and the student learning and teaching processes that take place within classrooms. Similarly, as more states embrace "charter schools," schemes that privatize schooling-related services (e.g., Edison, Educational Alternatives Inc.), or adopt school "choice," the educational landscape promises to become even more diverse. If public education remains a federal policy issue—and we think it must—then federal efforts that gather systematic, representative data on a myriad of schooling interventions seem more vital than ever. NCES has a key role to play in this effort.

NOTES

1. The concept of OTL was first introduced in the 1960s. For a full discussion of the origins, evolving definition, and policy applications of OTL, see McDonnell (1995) and other contributions to the special issue of *Educational Evaluation and Policy Analysis* dedicated to the late Leigh Burstein 17(3), Fall 1995.

2. The distinction is often drawn between the *intended* curriculum, the *implemented* curriculum, and the *attained* curriculum (Stecher 1992; McDonnell 1995). Here we are concerned with the first two of these.

3. The three reasons for collecting data on opportunity to learn distinguished below are similar to those outlined by Porter (1991).

4. For a review of process-product research, see Brophy and Good (1986) and Good (1995). Some recent work on curriculum and student achievement is also noted below.

5. For a review of “effective schools” literature, see Purkey and Smith (1983) and Rosenholtz (1985). For a useful critique see Rowan et al. (1983).

6. For a comprehensive overview of this work, see Hanushek (1986) and the critique by Hedges, Laine, and Greenwald (1994). Hanushek (1979) contains a useful discussion of methodological problems associated with this type of research. More recent examples using NCES data include Ehrenberg and Brewer (1994) (HS&B) and Ehrenberg, Goldhaber, and Brewer (1995) (NELS).

7. A recent example is Kuppermintz et al. (1995). While the primary purpose of this paper is to assess the validity of NELS mathematics test items, they do find strong effects of course program indicators on achievement. For example, higher scores in mathematics knowledge are associated with teacher reports of more traditional instruction methods (p. 545).

8. One shortcoming of surveys like NAVE for tracking reform concerns timing of data collection relative to the time it may take to implement reform efforts, especially curriculum or teaching practice reforms that can take many years to put into practice (Grubb and Stasz 1992).

9. For example, the 1965 Equality of Educational Opportunity Survey (Coleman Report) teacher survey contained several questions related to teachers’ qualifications and academic backgrounds, but very few directly to their classrooms (basically limited to number of students, hours spent preparing for class, number of classes taught, and attitudes toward racial issues and ability grouping). A recent re-examination of these data may be found in Ehrenberg and Brewer (1995).

10. For an overview of NCES surveys and future plans, see NCES (1995a).

11. While HS&B contains longitudinal student information, including standardized test scores at two points in time for the sophomore cohort of 1980, there is no direct link between the teachers surveyed and the students, or between the tests and particular classroom subject content. SASS includes no student component. NAEP, while containing directly linked student and teacher information on curriculum and instruction, and student standardized achievement measures, consists of representative samples of several grade/age levels; it is not longitudinal in the sense that the same students are surveyed from year to year.

12. Various “transcript studies” have been conducted in conjunction with HS&B (1982), NAEP (1987, 1990), and NELS (1992). These involve the coding and descriptive analysis of a large number of school transcripts obtained for a subsample of students in the relevant national survey (NCES 1995). They provide additional information about student course-taking patterns, though they do not indicate detailed curriculum content.

13. For details on future surveys, see NCES (1995a). The major scheduled vehicles for future data collection on curriculum and pedagogy at the K–12 will be NAEP (continuing every 2 years), and the Early Childhood Longitudinal Study (ECLS), planned to begin pilot testing in 1996–97. SASS (continuing every 5 years) will collect teacher data.

14. More details on these studies as they relate to “opportunity to learn” may be found in NCES (1995b).

REFERENCES

- Blank, R. K. 1993. "Developing a System of Education Indicators: Selecting, Implementing, and Reporting Indicators." *Educational Evaluation and Policy Analysis* 15 (1): 65–80.
- Blumenfeld, P. 1992. "The Task and the Teacher: Enhancing Student Thoughtfulness in Science." In J. Brophy (ed.), *Advances in Research on Teaching* 3: 81–114. Greenwich: JAI Press.
- Brophy, J., and Good, T. 1986. "Teacher Behavior and Student Achievement," In M. Whittrock (ed.), *Handbook of Research on Teaching*. New York: MacMillan.
- Burstein, L., McDonnell, L. M., Van Winkle, J., Ormseth, T., Mirocha, J., and Guitton, G. 1995. DRU-1086-NSF. *Validating National Curriculum Indicators*. Santa Monica, CA: RAND.
- Collins, A. In press. "Design Issues for Learning Environments." In S. Vosniadou, E. DeCorte, R. Glaser, and H. Mandl (eds.), *International Perspectives on the Psychological Foundations for Technology-Based Learning Environments*. Hillsdale, N.J.: Erlbaum.
- Collins, A., Seely, J., and Newman, S. 1989. "Cognitive Apprenticeship: Teaching the Crafts of Reading, Writing and Mathematics." In L. B. Resnick (ed.), *Knowing, Learning, and Instruction: Essays in Honor of Robert Glaser*. Hillsdale, N.J.: Erlbaum.
- Ehrenberg, R. G., and Brewer, D. J. 1994. "Do School and Teacher Characteristics Matter? Evidence from High School and Beyond." *Economics of Education Review* 13 (1): 1–17.
- Ehrenberg, R. G., and Brewer, D. J. 1995. "Did Teachers' Verbal Ability and Race Matter in the 1960s? *Coleman Revisited*." *Economics of Education Review* 14 (1): 1–23.
- Ehrenberg, R. G., Goldhaber, D. D., and Brewer, D. J. 1995. "Do Teachers' Race, Gender, and Ethnicity Matter? Evidence from NELS:88," *Industrial and Labor Relations Review* 48 (3): 547–561.
- Erickson, F., and Shultz, J. 1992. "Students' Experience of the Curriculum." In P. W. Jackson (ed.), *Handbook of Research on Curriculum*. New York: MacMillan.
- Good, T. L. 1994. "Teaching Effects and Teacher Evaluation." Draft manuscript of a chapter to appear in the second edition of the *Handbook of Research on Teacher Education*.
- Grubb, W. N., and Stasz, C. 1992. *Assessing the Integration of Academic and Vocational Education: Methods and Questions* (MDS-445). Berkeley: National Center for Research in Vocational Education.

- Guittton, G., and Oakes, J. 1995. "Opportunity to Learn and Conceptions of Educational Equality." *Educational Evaluation and Policy Analysis* 17 (3): 323–336.
- Hanushek, E. A. 1979. "Conceptual and Empirical Issues in the Estimation of Education Production Functions." *Journal of Human Resources* 14 (3): 351–388.
- Hanushek, E. A. 1986. "The Economics of Schooling: Production and Efficiency in the Public Schools." *Journal of Economic Literature* XXIV (3): 1141–78.
- Hedges L., Laine, R., and Greenwald, R. 1994. "A Meta Analysis of the Effects of Differential School Inputs on Student Outcomes." *Educational Researcher* 23 (3): 5–14.
- Hoffer, T. B., Rasinski, K. A., and Moore, W. 1995. *Social Background Differences in High School Mathematics and Science Coursetaking and Achievement* (NCES 95-206). Washington, D.C: U.S. Department of Education.
- Jones, L. V., Davenport, E. C., Bryson, A., Bekhuis, T., and Zwick, E. 1986. "Mathematics and Science Test Scores as Related to Courses Taken in High School and Other Factors." *Journal of Educational Measurement* 23 (3): 197–208.
- Kupermintz, H., Ennis, M. M., Hamilton, L. S., Talbert, J. E., and Snow, R. E. 1995. "Enhancing the Validity and Usefulness of Large-Scale Educational Assessments: NELS:88 Mathematics Achievement." *American Educational Research Journal* 32 (3): 525–554.
- Kozol, J. 1991. *Savage Inequalities: Children in America's Schools*. New York: Crown.
- McDonnell, L. M. 1995. "Opportunity-to-Learn as a Research Concept and a Policy Instrument." *Educational Evaluation and Policy Analysis* 7 (3): 305–322.
- Monk, D. H. 1992. "Educational Productivity Research: An Update and Assessment of its Role in Education Finance Reform." *Educational Evaluation and Policy Analysis* 14: 307–332.
- Murnane, R. J., and Phillips, B. 1981. "What do Effective Teachers of Inner-City Children Have in Common?" *Social Science Research* 10: 83–100.
- National Center for Education Statistics (NCES). 1995a. *Programs and Plans of the National Center for Education Statistics*. Washington, D.C.: U.S. Department of Education.
- National Center for Education Statistics (NCES). 1995b. *Measuring Instruction, Curriculum Content, and Instructional Resources: The Status of Recent Work* (NCES Working Paper 95–11). Washington, D.C.: U.S. Department of Education.
- National Council of Teachers of Mathematics (NCTM). 1991. *Professional Standards for Teaching Mathematics*. Reston, VA: NCTM.

- Oakes J., Selvin, M., Karoly, L., and Guitton, G. 1992. *Educational Matchmaking: Academic and Vocational Tracking in Comprehensive High Schools* (MDS-127). Berkeley: National Center for Research in Vocational Education.
- O'Day, J., and Smith, M. S. 1993. "Systemic Reform and Educational Opportunity." In Susan Fuhrman (ed.), *Designing Coherent Education Policy*. San Francisco: Jossey-Bass.
- Porter, A. C. 1991. "Creating a System of School Process Indicators." *Educational Evaluation and Policy Analysis* 13 (1): 13–29.
- Porter, A. C. 1995. "The Uses and Misuses of Opportunity to Learn Standards," *Educational Researcher* 24 (1): 21–27.
- Purkey, S. C., and Smith, M. S. 1985. "Effective Schools: A Review." *The Elementary School Journal* 83 (4): 427–452.
- Reigeluth, C. M. 1987. *Instructional Theories in Action: Lessons Illustrating Selected Theories and Models*. Hillsdale, N.J.: Erlbaum.
- Rowan, B., Bossert, S., and Dwyer, D. 1983. "Research on Effective Schools: A Cautionary Note." *Educational Researcher* 4: 21–31.
- Serrano v. Priest*. 1971. 487 P.2d 1241. "Serrano v. Priest: Implications for Educational Equality." *Harvard Educational Review* 41 (4): 501–534.
- Shavelson, R., McDonnell, L. M., Oakes, J., and Carey, N. 1987. *Indicator Systems for Monitoring Mathematics and Science Education* (R-3570-NSF). Santa Monica: RAND.
- Smithson, J. L., Porter, A. C., and Blank, R. K. 1995. *Describing the Enacted Curriculum: Development and Dissemination of Opportunity to Learn Indicators in Science Education*. Washington, D.C.: Council of Chief State School Officers.
- Stasz, C., McArthur, D., Lewis, M., and Ramsey, K. 1990. *Teaching and Learning Generic Skills for the Workplace* (R-4004-NCRVE/UCB). Santa Monica: RAND.
- Stasz, C., Ramsey, K., Eden, R., DeVanzo, J., Farris, H., and Lewis, M. 1993. *Classrooms That Work: Teaching Generic Skills in Academic and Vocational Settings* (MR-169-NCRVE/UCB). Santa Monica: RAND.
- Stecher, B. 1992. *Describing Secondary Curriculum in Mathematics and Science: Current Status and Future Indicators* (N-3406-NSF). Santa Monica: RAND.
- Stigler, J., and Fernandez, K. 1995. "Videotape Classroom Study: Field Test Report" (unpublished technical report). Los Angeles: University of California.

Teacher Education, Training, and Staff Development: Implications for National Surveys

David R. Mandel

INTRODUCTION

The national conversation about teaching has always been compromised by a dearth of information about the quality of practice and practitioners. Without such information, policymakers and the public have been left to fend for themselves, often speculating wildly about the current state of teacher education and teaching. When dismal or promising results about student performance are reported, a new chain reaction of suppositions is often set off about the degree to which teachers are to be blamed or praised. But these suppositions are just that—hypotheses disconnected from much of a factual base that might shed some light on what is occurring, including the extent to which the observed results can be accurately attributed to teacher actions.

All of this should not be the least bit surprising, because until recently the profession had not been able to agree on just what are the essential elements of highly accomplished practice. Without such a definition, those charged with collecting data on teacher quality were left in the lurch. As a result, to the extent that any qualitative measures of teacher competence exist, they are marginal at best. This dilemma also hobbles the necessary debate that should be ongoing about teaching, and it compromises the ability to conduct useful research and analyses about teacher, school, and system performance. These circumstances have led to the growth of some familiar myths about teaching, such as all that matters is teachers' command of subject matter knowledge; teachers just need to stay one chapter ahead of students and employ the latest pedagogical trick; or anyone who is a warm, nurturing, and caring individual can be a teacher. By abdicating its responsibility to put forth a rich description of excellent practice, the profession has contributed to the promulgation of these notions that teaching is for amateurs, not professionals, and that anyone with a good heart and a modicum of intelligence stands a decent chance of being successful.

A UNIQUE OPPORTUNITY TO ADDRESS THE TEACHER QUALITY ISSUE

Fortunately, the nation is moving to a new reality as, for the first time, a systematic and prolonged effort is being undertaken to develop in each field of teaching a professional consensus about what accomplished teachers should know and be able to do. Constructed as part of a larger

effort by the National Board for Professional Teaching Standards (NBPTS) to create a voluntary system of advanced professional certification for early childhood, elementary, middle, and secondary school teachers, this new reality has emerged in the form of standards for advanced practice that provide a foundation for important conversations about teacher quality that previously could not be conducted (NBPTS 1994, 1995, and 1996).

The National Board standards, focused as they are on the critical aspects of teaching that distinguish the practice of accomplished professionals, create a framework for evaluating teachers, institutions, and programs that are designed to develop exemplary teachers. Such a framework heretofore could not be provided by teacher licensing standards, for important as they continue to be to the states (as they exercise their responsibility to assure basic competence in individuals assigned responsibility to educate the young and vulnerable), they are constrained by the function they are designed to serve. Concurrently, these standards for exemplary practice allow the profession to claim the legitimacy it is due, because they define the expertise and best practices that not only distinguish highly accomplished practitioners from beginners and journeymen, but also make clear to the public as well as professionals the elegance that marks excellent practice when the art and science of teaching are well joined.

The National Board's vision of accomplished teaching recognizes that the world for which we are preparing our students is importantly different from the mission schools have focused on in the past. Like the best of the recent student standards initiatives, this view of teaching is built on a common ground of assumptions that begin with an acknowledgment that the nation can no longer afford to provide an excellent education to a small elite and a pedestrian education to the masses. The reasons are not only our changing economic circumstances (which are quite significant in and of themselves), but also the need to have a much better educated citizenry to sustain the nation's democratic values and institutions. These standards efforts also proceed from the shared view that neither civic competence nor labor market success can be achieved by simply stuffing students full of facts, rules, and theorems. While such a knowledge base is important, it must be augmented with the ideas, concepts, theories, and knowledge of the core disciplines that allow the well-educated to constructively address the challenges that face us daily at home, in our communities, and in the workplace. The ideas that teaching should be more than telling, that students should be actively engaged in learning and applying knowledge, that in-depth understanding is to be valued over coverage, and that all children can learn are at the heart of this perspective.

This new and emerging agreement on first principles captured in the policies and standards of the National Board for Professional Teaching Standards (NBPTS 1991) provides an opening for the National Center for Education Statistics (NCES) to aggressively pursue a set of issues that have interested it for some time, but up to now have proven elusive. As one reviews report after report about teaching that NCES has commissioned in the aftermath of *A Nation At Risk*, one is struck with the regularity with which the authors stub their collective toes on the teacher quality issue—not for lack of trying, but because there has been little data with which to work.

In the past several years, the Center has made important strides in uncovering the complicated world of teacher migration, helping the policy community better appreciate the intricacies of teacher movements from position to position within the system and of entry into and exit from the system. This has illuminated the discussion of teacher supply and demand, probably shattered some old

myths that needed to be set aside, and made the discussion of this subject much more sophisticated and valuable. Still, having a better grasp of the raw flows of teachers (potential, current, and former) in and around schools provides modest solace when there remains little good understanding of the qualitative dimensions of such flows (e.g., are the teachers leaving the work force stronger than those staying, or vice versa). If there is an increase or decrease in new teachers entering or mature teachers leaving the profession should the nation be concerned? The answer depends on the mix of teachers involved and the quality of the stream of those entering, migrating, and leaving—on this score parents, policymakers, and administrators are largely in the dark.

What is known is the type of education credentials teachers have accumulated and the type of state licenses they have been granted. This information has proven useful in gaining a rough sense of how well-prepared teachers are to take on the assignments they are handed; thus, the public has been alerted to the high percentage of students whose education has been entrusted to individuals who do not have a basic grounding in the subject(s) they teach. But such data, even when positive, provide only the most modest threshold of confidence regarding the quality of practice in the nation's schools. This is so for several reasons, not the least of which is that licensure requirements differ markedly from state to state,¹ and the fact of completing a major or minor in a subject holds substantively different degrees of meaning from campus to campus. However, even if these issues of non-uniformity could be swept off the table in a stroke, yet another fundamental problem remains: assurances of minimal competence tell us hardly anything about the quality of practice once teachers have moved beyond their first few years of practice. This is a field where expertise takes time to develop and where exemplary teachers build a repertoire of content-specific pedagogy over time on which they can draw, refine their ability to understand what their students grasp about important matters, and become more astute in making the sound and principled professional judgments about how best to proceed on a daily basis—decisions that are crucial if all students are to be presented with the requisite opportunities to learn to their fullest. In short, only the first phase in educating and developing accomplished practitioners has been completed when a first degree is awarded and a first license granted. Thus, if the education community and the public at large are to have a genuine sense of the quality of the teacher work force, data must be collected on a regular basis from across the profession that are representative of the various career paths and education and work regimens that teachers have experienced and that have a profound influence on the quality of teaching in America's schools. Simply put, we need to move toward gathering information on what teachers know and can do today, and to adopt an analytical stance that is less fixated on the knowledge, skills, and dispositions they brought with them on their first day of teaching.

THE CENTRALITY OF INDICATORS OF TEACHER QUALITY TO EDUCATIONAL POLICY

This paper focuses on teacher quality because it is the quicksand that mires almost all large-scale data analysis of the state of teaching, teacher education, training, and staff development. Although NCES collects useful information on many characteristics of the nation's teachers and the type of educational experiences to which they have been exposed, the value of such data is diluted and the investment in its collection marginalized when the necessary qualitative information that ought to be linked to these other characteristics is absent. The result is that with only the most

perfunctory descriptive information collected and reported, policymakers can and do fail to take critical actions that they might otherwise take to bolster the quality of education our children receive, or act precipitously when such actions are not warranted, or both.

The large policy issues that should drive data collection and analyses in this arena can be captured in the following quality assurance themes:

- The overall quality of the teaching force;
- The quality of the teacher education system; and
- Trends in teacher quality, institutional supports, and incentives for improvement.

The Overall Quality of the Teaching Force

The bedrock of the nation's schools are the teachers who are its front-line workers. The schools cannot be any better than their teachers, who, if they are to move students to high levels of performance, must make careful professional judgments on a daily basis about how students spend their time and on what subjects, ideas and concepts they focus their attention. Teachers create or fail to create learning environments that motivate student effort and stress democratic values. They teach or fail to teach the perspectives, dispositions, forms of inquiry, and ways of knowing that mark the core disciplines and that provide the pathways to the important ideas that students need to grasp. They may increase to some greater or lesser degree students' ability to learn on their own, to work collaboratively with others, and to develop the values, character, and knowledge that will allow them to function well as adults in the marketplace and in their communities and its democratic institutions. Just what the capacity of the teaching force is to perform these functions and the extent to which they are performed well is crucial information that is desperately needed from both a public policy and public administration perspective. Data are needed not only on an aggregate basis but also with respect to particular teaching fields and locales. Without such knowledge of the teaching force, those seeking to improve the schools to meet the very real challenges this society faces both externally and internally are operating with one hand tied behind their back, ignorant of the depth and breadth of the problems they confront or the assets they possess at a key point of leverage in the system.

The Quality of the Teacher Education System

A key determinant of the quality of teaching is the kind and quality of investments that are made in the initial and continuing education of teachers throughout their careers. Here, too, there is the thinnest veneer of hard evidence on which to draw conclusions, yet one can find both an extraordinary degree of complacency in some quarters about the state of this substantial enterprise and calls for the complete dismantling of the system in others. That both views have legitimacy concurrently is quite stunning. It is understandable, however, once one recognizes that in the absence of reliable and trustworthy measures (e.g., a national accreditation system with teeth to which all states subscribed, or a teacher education "NAEP") to provide a bulwark against unsubstantiated claims, such claims, no matter what distance from reality, can gain currency.

This circumstance is highly problematic in the current policy environment where for the past decade teacher education, be it undergraduate, graduate, or continuing professional education, has been under attack by its clients, the policy community, and many of its most distinguished members. Some of the concerns appear ill-founded, whereas others appear well-grounded. While it is true that accreditation by the National Council for the Accreditation of Teacher Education (NCATE) provides an increasing measure of quality assurance for those schools of education that do open themselves to critical examination by their peers, most do not. Furthermore, no parallel check on quality exists for the millions of dollars sunk into the ongoing professional development of teachers that occurs outside NCATE institutions. Even within NCATE schools, all that is known is that a basic threshold of quality has been met. The degree to which these institutions in general or the particular programs that they operate exceed the threshold remains a mystery.

Trends in Teacher Quality, Institutional Supports, and Incentives for Improvement

Although educational institutions are often characterized by their insularity to externally imposed change and by a high level of inertia even in response to internally provoked change, they have their dynamic aspects. Similarly, the teaching force itself rarely remains static for long, as shifts in the demography of the country and in the particulars of the labor market inexorably make their presence felt over time. Even without such forces at play, changes in the number of students, pupil-teacher ratios, compensation, and other conditions of work often yield effects of similar character. These shifts may be unevenly felt in different regions of the country and in different teaching fields, but they are unavoidable. NCES has played a valuable role in helping the education system gauge well the nature of such trends in the past and should continue to do so in the future.

To date, however, the focus has been on large flows through the system without much attention to quality for all of the reasons cited above. With the addition of some qualitative dimension to such analyses much more informative work could be conducted. More robust data on teacher quality would also allow analysts to look at the effects of various incentives and programmatic initiatives to improve teaching that are now, for the most part, elusive.

GOVERNING PRINCIPLES

In thinking through a new strategy for NCES to consider in this arena, a few key maxims might serve to guide the conversation. For example:

- “You can’t boil the ocean.” This rule, borrowed from Lewis Branscomb when he was Chief Scientist at IBM in the 1980s, is the obverse of the “there is no silver bullet” theory. It warns that agencies need to think hard about what investments are likely to have the greatest payoffs, as it is too easy to fall into the trap of trying to address every known issue to the detriment of addressing a few critical ones well.
- The easiest data to collect may not be especially valuable. Knowing how many teachers will retire this year has modest value until that information can be placed in some larger context and some further information gathered about just whom is retiring (e.g., what they teach and whether or not they are among the strongest teachers in the system).

- Some cheap proxies may be worth looking at, either because they are cheap or because NCES has a long history of collecting them and their behavior is telling in one way or another.
- Find out what the customers value (in the current plan and prospectively) by talking to federal, state, and local policymakers; teacher educators; the unions; disciplinary and specialty associations; and the research community.
- Qualitative indicators in this field that are likely to be trustworthy are not going to emerge from counting anything (at least for the next several years, at which time the percentage of National Board Certified Teachers in a jurisdiction will mean something). Such indicators depend on the professional judgments of well-educated and well-trained examiners/observers.

CRITICAL PERSPECTIVES

In thinking through an agenda for future NCES initiatives that are focused on teachers and teaching, the history and policy questions sketched out above provide the backdrop for decision making. In considering the range of options that seem most promising to explore in light of these factors, four core frames of reference are suggested:

- *Qualitative measures of individuals*—which are without question the most important measures to capture;
- *Qualitative measures of institutions*—which pose substantial conceptual, economic, and political challenges, but would represent a breakthrough of significant proportions in understanding teaching at the postsecondary level;
- *Flows through the system*—with special attention to activities that hold the potential to promote the upgrading of the teaching force and minority participation; and
- *Kind and level of investment in teacher quality*—in time and fiscal resources by all the various players, including the teachers themselves.

A discussion of each core frame of reference follows.

Qualitative Measures of Individuals

Coming to grips with measuring the quality of teachers is not going to be addressed with a better survey instrument, even if it had an examination imbedded within it. This is because what teachers seem to know and can write about is not always an accurate predictor of what actually transpires in their classrooms. This truism has been rediscovered anew by researchers visiting the classrooms of mathematics teachers who claimed they were implementing the new National Council

of Mathematics Teachers' standards. While many of these teachers spoke convincingly about how they had adapted their practice to the new standards, the actuality was often significantly different.

The hard reality is that there is no substitute for actually observing teachers at work with students. This observation could take a variety of forms, including the use of video, but it cannot be approximated by absenting the observer from the hurly-burly of the classroom. This fixation with observing the act of teaching as a necessary prerequisite to judge teacher quality is not to discount the value of NCES also providing accurate data on teachers' knowledge of the subjects they teach, of the content-specific pedagogy associated with these subjects, and of the misconceptions and difficulties about important ideas that are common among students in the age range they are teaching. As NCES proceeds down this path, it should strive to capture both what teachers know and what they can do in terms of the profession's highest standards, those established for National Board Certification in each field of teaching.

This is a tall order and presents many logistical, administrative, and economic hurdles, but with some imagination they ought to be surmountable. For starters, NCES should consider sampling teaching in a few core fields (e.g., primary grades instruction, middle grades English, high school mathematics) once every 5 years, thus providing the nation with the equivalent of a NAEP for teachers. Gaining a sense of teachers' command of the knowledge base of their field would require NCES administering some form of written examination to a sample of the teaching force, a nontrivial exercise in political and economic terms until such time as the states begin to require concrete evidence that teachers are keeping abreast of new developments in their field as part of state relicensing requirements.

More plausible in the near term would be an effort to ascertain the nature and quality of actual teaching practice across the land. To do so, the Center could take advantage of the assessment technology now being developed by NBPTS, which joins videos of teachers' practice with student work samples and teacher commentaries on the goals of instruction, the context in which the instruction is being conducted, the quality of work and understanding of the students, and teachers' rationale for proceeding as they do. These videos and their surrounding artifacts and explanations are each focused on a central responsibility that teachers in each field must discharge to advance student learning (e.g., in English language arts there is a video exercise on teachers working with students to interpret one or more texts the students had read). A 15- to 20-minute, continuous tape of such teacher/student interactions is generally quite revealing.

Viewing teachers' actions in this manner and joining such evidence with teachers' justifications for their actions and interpretations of student performance can prove especially telling. There will be questions about how much evidence is enough; about how much burden can be placed on teachers; and about the trade-offs in validity, reliability, and administrative feasibility between this proposal and some form of on-site observation. Both approaches deserve consideration, as does the question of paying teachers to assemble a mini-portfolio versus paying observers from outside their district to conduct on-site observations. Each approach has advantages and disadvantages. NBPTS has opted for the former because it allows the National Board to collect multiple samples of practice at low cost, avoids the idiosyncracies of schooling that can disrupt the class on any particular day, and creates a permanent record of teaching that can be referred to as many times as necessary if different judges take markedly different views of the practice or if, at a later time, there

are legitimate grounds for an appeal.² However, the purposes of NCES and the National Board in collecting such data are significantly different and, consequently, replicating the NBPTS methodology may not necessarily be the best course.

Whatever method of observing practice is chosen, over the next several years the NBPTS system for advanced certification will produce a growing cadre of National Board Certified Teachers in multiple fields, some number of whom have been trained to score such evidence produced by peers and have also been found to be fair and reliable judges. The presence of this cadre provides NCES with a head start in being able to evaluate the data the Center could collect on teacher quality.

Moving forward in this direction would mark the first steps in developing a regular report to the nation on *The Condition of Teaching*. Knowing far more about the quality of teaching in particular fields (e.g., which aspects of science teaching teachers seem to have good command over and which they do not) would provide valuable guidance about where resources for teacher education need to be directed, not only within particular fields but also between fields as well. Understanding teachers' command of subject matter as well as the extent to which their beliefs about pedagogy conform to the current professional consensus in their field would be especially illuminating. Such data might move preservice programs to rethink some of their underlying assumptions, alert teachers and administrators to areas of potential weakness, and cause school districts to better target their scarce resources on professional development initiatives where the need is greatest.

The trap door to avoid in conducting such work is the trivializing of the complexities of teaching that can occur in a mad rush to design super efficient instruments. NBPTS has worked this territory hard to create professionally acceptable and administratively feasible assessments. On this score, the National Board has learned several lessons that should serve NCES well, including the need to avoid overly atomistic measures of practice that artificially disassemble the components of practice to the point where they lose their meaning. Recognizing that for many instructional issues there are multiple sound approaches, that judgments about the efficacy of teachers' actions need to take account of the instructional context, and that maintaining the authenticity of practice is especially important are other key considerations that must be attended to in such work.

Qualitative Measures of Institutions

However formidable these ideas for taking periodic measures of the quality of teaching may seem, they pale in comparison with designing a parallel plan for institutions and organizations that supply teacher education, which by virtue of performing this function play a significant role in shaping the kind and quality of teaching found in the schools. While attempting to develop such institutional measures may appear to be a fool's errand in light of the allergic reaction demonstrated by many institutions to any serious form of quality assurance (one need only have observed the withdrawal of colleges and universities from NCATE when faced with the reality or the potential of a negative finding), the uneven quality of teacher education that careful observers report they find nationwide suggests that progress on this front could have substantial benefits. The large number of institutions providing teacher education also argues for serious attention to this issue, as the

proliferation of programs has raised well-founded concerns that weak programs soak up resources that might better be concentrated on upgrading those with greater potential for quality instruction; and in the bargain, the overall quality of education that is provided to prospective teachers is being compromised. The idea would not be to rank or rate individual institutions, but to provide a portrait with a good deal of fidelity that would accurately reflect the current state of teacher education.

What might such an effort be? At a minimum, it should consist of a set of careful case studies that NCES would regularly mount that would yield thoroughgoing portraits of the range of programmatic approaches being undertaken not only in preservice education but also in continuing education.³ The latter arena is extraordinarily fractionated as post-licensing education takes place in traditional college and university settings, as well as in teacher centers, in school districts, in seminars and courses run by disciplinary and specialty groups, and in other informal settings. But this diversity of approaches should not discourage NCES, because it is in just such settings that some of the best and some of the worst teacher education is located. In addition, it is in the quality of such post-licensing education that the prospects for a novice teacher to 1-day advance to expert status are forged. So it is an arena that is crucial to the health of the profession and deserves much more attention than it has traditionally been accorded.

NCES should also consider moving beyond the case study option to a format that more closely parallels the studies of teaching quality suggested above. While this approach would represent a sharp departure from current practice, it should not be dismissed out-of-hand because its potential to illuminate this critical aspect of teacher education is substantial.

What better way to judge the quality of teacher education and disarm those critics who claim it leads nowhere than to have NCES engage an independent auditor with some stature to visit a good cross-section of the nation's colleges of education and look beyond the plans, the curricular offerings, the admissions and graduation requirements, library holdings and faculty credentials to the actual teaching of prospective teachers that takes place? The objective would be to paint a picture of the kind and quality of preparation and ongoing professional development that the nation's current and prospective teachers are experiencing whether in schools of education or in colleges of arts and sciences.

To conduct such work, standards of teaching teachers would have to be developed, and this task should not be conducted by NCES but by some independent entity with fiscal support from NCES.⁴ This exercise in and of itself would be healthy for the entire higher education community, as it would establish a set of commonly shared principles that would frame a host of instructional and curricular decisions that college and university faculty have to make on a regular basis. It would also provide guidance for designers and sponsors of continuing professional education and for consumers of their services.

Establishing standards for teaching teachers would also be especially timely given the efforts of several reform initiatives to provide the enterprise with substantial uplift by placing greater emphasis on clinical training, on the closer integration of subject matter and pedagogy, on developing the habit of self-examination, and on the translation of theory into practice through professional development schools (i.e., the educational equivalent of teaching hospitals) and other vehicles.

Flows Through the System

As student enrollments have ebbed and flowed, the demographic turmoil that was periodically provoked in the schools led inexorably to a growing interest in data on teacher supply and demand. However, in an environment where teacher quality is not an issue, the supply of teachers is inexhaustible. Still, the devotion of NCES to collecting data on the movement of teachers within the system has been rewarded with findings about how minimal standards can be manipulated in the face of potential teacher shortages—an underreported scandal in American education. Whether it is the issuance of emergency licenses, the lowering of entry-level standards of competence, or the misassignment of teachers (an example of a low-cost proxy for the larger qualitative problems that plague teaching), such threats to the quality of education America's children receive need to be closely monitored.

Improvements in the ability of NCES to understand the complexity of teacher flows both within and in and out of the teacher work force should now permit finer grain studies of change over time and the effects of changes in policy on such flows. Of special interest on this score is the degree to which the professionalization of teaching may affect who is attracted to teaching, who stays, and who leaves. Today, the incentive structures in most local systems encourage teachers to become administrators, resulting in the loss of some of the strongest practitioners from the classroom. National Board Certification and other worthwhile initiatives are making a concerted effort to change this reality and to change the culture of teaching to provide better support to novice and experienced teachers, to value continuous learning and problem solving in the company of one's peers, to recognize and reward excellence, and to alter schools' organizational structures to capitalize on the knowledge, skill, and expertise of the profession's strongest practitioners. If these efforts begin to take hold, teaching should become a much more attractive career option, and its ability to attract and hold talented people with many other career choices should improve.

To the extent these changes materialize on a large scale, they will do so gradually, and they will most likely occur unevenly across the education landscape, as risk taking of this sort is not the common condition usually found in the schools. Thus, it will be important not just to be able to track the aggregate flows of teachers in and around the system, but also to look for changes in patterns linked to changes in policies of the sort noted above. In the short term, this may mean more targeted studies focused on those states and localities where these initiatives are beginning to take off.

In the various efforts that NCES may undertake to track teachers, special attention should be given to the dynamics of minority teacher movements, because there remains a sharp disjuncture between the percentage of the teaching force that is drawn from the minority community and the percentage of minority students in the schools and the percentage of people of color in the labor force. This issue needs attention not because minority children must be taught by minority teachers, but because the teaching force itself would be strengthened by teachers with more diverse life experiences and all students would benefit from working with a range of adults who are more reflective of the larger society they are being prepared to enter. It is important to understand how the career paths of teachers of color and those of their white counterparts are similar or different, and to understand to what extent these differences are either being magnified or dampened over time.

All such work will be markedly enhanced as the numbers of National Board Certified Teachers (NBCTs) begin to multiply, because it will allow analysts to bring a previously unattainable qualitative dimension to their studies. As NCES thinks about future requirements, planning to track the movements of NBCTs should be a high priority. The kind of education they have had throughout their careers, the nature of their teaching experiences, the conditions of their workplaces, and the changes in their employment circumstances over time should prove especially illuminating. By following this course, clues to the merits of various teacher recruitment, hiring, assignment, and compensation policies would be subject to more searching investigations, and the question of the uneven distribution of teaching talent could be given the concerted attention it is due. On this latter point, it is well understood by those close to the system that inequities in school finance and other factors result in some schools having their pick of exemplary teachers whenever a vacancy occurs, while others serve as farm teams for their more advantaged colleagues. Lacking valid measures of teacher quality, such discrepancies have been too easy to overlook. With the advent of National Board Certification not only can sound analyses be conducted on this point, but also the effects of new policies to ameliorate this problem can be reasonably judged.

In time, one would also want to examine the effect of NBCTs on the practice of those who work in close proximity to them, but such studies probably are best sponsored by the research arm of OERI.

Kind and Level of Investment in Teacher Quality

Although the nation is fortunate to have many excellent teachers in its schools, there is no question that it needs more and that the overall quality of teaching needs to improve significantly. This will not happen overnight, nor will it happen without a concerted effort by many parties. In fact, it will require a sea change in labor-management relations, a break from past practices along the entire front of teacher education, and a willingness by states and localities to commit to a dramatic reconception of how to foster the growth of exemplary practitioners. This means that schools will have to find better ways to invest the resources they currently allocate to teacher education (not treating it as a fringe benefit and not just paying for the accumulation of graduate credits at the cheapest and most convenient institution teachers can find), while simultaneously increasing their investment in their most valuable capital assets, teachers.

Just how school districts take up this challenge is crucial, and the extent to which teachers meet their employers halfway is no less important. NCES can contribute to this effort by monitoring it closely through the Schools and Staffing Survey or through special supplements to it. Some of the more interesting questions on the table where data collection and analyses could be especially telling are the following:

- What are teachers doing to strengthen their practice? How much effort are they expending in this direction? How much support/encouragement are they receiving in this direction, and what form is it taking?

- What is the level of public sector investment in initial teacher preparation and ongoing professional development? Where are these investments being made, and what are they purchasing?
- What is the level of public and private sector investment in research and development to improve teaching?

Pursuing these questions and others that they generate should provide NCES with a rich but manageable challenge. Unlike the dicier questions of measuring teacher quality discussed above, these matters should yield to the more common methodologies at which NCES is well practiced.

EPILOGUE

Education has been through a period where it has been subject to high-level, high-intensity, and high-volume scrutiny. While it has been painful for some, it is an extraordinarily promising exercise for the country because it is a sign that the nation understands that excellence and equity in education are necessary prerequisites to a healthy society. As this debate has proceeded, there appears to be a growing recognition of the centrality of the quality of teaching and of the ways in which the education and training of teachers proceeds. Surprising as it may seem today, this was not the focus of many early efforts at education reform in the 1980s, and it remains overlooked by some to their detriment even in the 1990s. Still, as this paper is written, we are just beginning to get our feet wet in this crucial arena of teacher quality.

With the advent of more valid measures of teacher knowledge and performance now emerging from the profession, NCES is well positioned to harvest the work of those toiling in the teacher quality vineyard and to begin new streams of data collection that will take advantage of these path-breaking efforts. This is work that a federal agency could not have conducted on its own without encountering substantial legitimate opposition. But now that others have taken the lead in defining the parameters of highly accomplished teaching, there exists a special opportunity to advance the work of NCES that should not be missed. And, that is to join in pushing the frontiers of this field of education measurement forward by investing in valid, reliable, and efficient means of gathering information on the quality of teaching and on the quality of the system support structures that can contribute to developing the excellent teachers America's schools desperately need in much greater numbers.

One should acknowledge that much of what is proposed in this paper is a significant departure from standard operating procedure at NCES, but the unavoidable fact is that these procedures in many respects are extraordinarily limiting. If NCES is only to continue along the current track it is following, it will be pursuing a strategy that will not yield much of the data the country requires to address some of its most difficult education problems. However, expanding its vision in some of the ways suggested here, or in other ways but with the same objectives in mind, should increase the prospects that key public policy decisions will be grounded in hard won intelligence about the system's characteristics and be less subject to influence by ignorance, guesswork, or untested theoretical constructs. If such a result ensues, American schools and students will be the beneficiaries.

NOTES

1. For example, only a few states have begun to insist on observing novice teachers during their early years of practice as part of their determination of competence before awarding a full-fledged license. Thus, being able to assure the public that teachers are able to convert theory and knowledge into sound practice in the early years of their practice is a warrant that accompanies only a modest fraction of the state licenses in circulation today.

2. Teachers also find this process commendable on several grounds. They get to put their best foot forward. They find it a means to open a new professional conversation with colleagues about the trials, tribulations, and achievements of their practice. And, they find the process fair, in part, because all of the judgments are made by peers from outside their school district.

3. This proposal would clearly be a departure for NCES, as such work has historically been sponsored by other OERI entities rather than the one charged with collecting, analyzing, and interpreting statistical data.

4. This is a model similar to that invoked to provide federal support for the discipline-based student standards and standards for advanced certification of elementary and secondary school teachers.

REFERENCES

- National Board for Professional Teaching Standards. 1994. *Early Adolescence/English Language Arts Standards for National Board Certification*. Detroit, MI.
- National Board for Professional Teaching Standards. 1994. *Early Adolescence/Generalist Standards for National Board Certification*. Detroit, MI.
- National Board for Professional Teaching Standards. 1995. *Early Childhood/Generalist Standards for National Board Certification*. Detroit, MI.
- National Board for Professional Teaching Standards. 1996. *Middle Childhood/Generalist Standards for National Board Certification*. Detroit, MI.
- National Board for Professional Teaching Standards. 1991. Chapter II: *What Teachers Should Know and Be Able to Do. Toward High and Rigorous Standards for the Teaching Profession: Initial Policies and Perspectives of the National Board for Professional Teaching Standards*. Third Edition. Detroit, MI.

Discussant Comments

MICHAEL TIMPANE

We are at a time when education policy presumes to say that it will change students' achievement outcomes measurably, either through systemic policies that clearly link standards, curricula, school programs, pedagogy, and assessments, or through the interplay of market forces in various systems of school choices. It is also a time when researchers are painstakingly establishing the combinations of content, pedagogy, and learning that characterize particularly successful classrooms and schools. In these circumstances, it is of the greatest importance that policymakers and program planners have reliable data on patterns of classroom activity and teacher performance, to know where we started, how things change, and with what effect. Thus, NCES should accord a very high priority to pursuing the lines of development suggested by Brewer and Stasz and by Mandel.

That said, the authors, correctly ambitious in laying out the possibilities, are equally correctly ambivalent about the prospects for success. They are correct to wonder how much of the needed data could be collected by survey instruments, however subtle and sophisticated they may be; the examples they present from recent studies are as halting as they are hopeful in sketching the pace of progress. All in all, their papers raise a formidable list of conceptual and empirical concerns:

- How can such data collections reliably connect teaching practices with levels or patterns of learning? And how should such learning be defined and measured?
- To what extent must such surveys be preceded and/or complemented by other small-scale studies involving interviews with teachers, classroom observations, review of logs, artifacts performance assessments, and so on?
- How would data collection and analysis take account of explanatory policy variables, such as teacher preparation and rewards, incentives and capacities for change, legislative and regulatory constraints ranging from teacher certification to union contracts to federal and state program requirements?

Staggering though the designs problems may be, let the predictably lengthy development process begin. We must have more compelling information about the ways in which our expectations—as set forth in policies and paradigms of teaching and learning, backed up with resources—are realized in schools and classrooms, and with what effect. Otherwise, our research and data collection will seem increasingly marginal to educational policy and practice, and educational policy and practice may seem that much more marginal to our nation's progress.

EILEEN M. SCLAN

Both the Brewer and Stasz and the Mandel papers suggest that NCES look at the multidimensional nature of school processes and how interactional effects of the teaching and learning process either enhance or reduce opportunities to learn. I support this effort. In this response, I will focus on two main issues: 1) the importance of contextual data and 2) its relevance to equity issues. Collecting contextual information about curriculum, student learning processes, instructional resources, and professional development of teachers requires a new way of thinking about data collection in schools. The challenge for NCES is to move beyond gathering data that reflect a static view of teachers' and students' realities toward collecting data that captures a dynamic view and that accounts for contextual variation and the vital nature of the teaching and learning processes. To begin the unprecedented task of gathering rich, detailed data on what goes on in schools and classrooms will require using non-survey methodologies.

Because teaching requires active involvement and is not something that is done to students, data are needed on student characteristics, the teaching and learning process, and teacher-student interactions to provide useful information to researchers, policymakers, and teachers themselves in understanding teaching and learning in its richer context. What students think, what they feel, and what they experience in and out of school influence the way they learn and how well they learn. In response to this need, innovative teachers have begun to implement more authentic assessment measures to better understand their students' learning. Both papers discuss the importance of employing authentic methods of assessing student and teacher performance through the following methods: teachers' work samples; logs; teachers' judgments in individual situations; the context of the teaching situation; interviews; observations; videotaping; and artifacts such as homework assignments, quizzes, texts, and exams. Qualitative methodologies may provide a closer look at how teachers can nurture students in becoming active, responsible, and responsive decisionmakers.

Workplace structures also influence the prevalence and nature of teacher and student opportunities to learn. In Rosenholtz's (1989) seminal work on teachers' workplaces, for example, the extent of teachers' learning opportunities was associated with increased student achievement gains. The nature of the social structures, policies, and traditions of the school environment often determine teachers' opportunities to learn. Teachers who have opportunities themselves to stay abreast of the exploding knowledge base in developmental psychology, social organizational theory, state-of-the-art teaching techniques, and subject matter content are more likely to provide opportunities for students that enable them to think critically, creatively, and deeply. Mandel aptly points out that investments in ongoing teacher education are key determinants of the quality of teaching. Teachers say they need institutional supports, such as time to meet with colleagues and to participate in policymaking at the school or district level, access to instructional materials and to recent research journals, and opportunities to attend professional conferences. To understand schools, we must understand them as teachers and students do. The authors call for a greater focus on the dynamic social organizational dimensions of workplace incentives that support teachers in becoming more effective, reflective, and analytical practitioners.

Contextual data on the wide spectrum of student experiences in classrooms, the nature of their opportunities to learn, and how they process particular forms of knowledge delivered in varying ways will provide us with windows to view the complexities of teaching and learning.

This information will allow us to study the more subtle interactional effects between teachers and students and workplace environmental variables, and how these factors, in turn, contribute to teaching for higher level problem solving, analyzing, critical and creative thinking, and deeper levels of appreciation for subject matter, issues, and people.

Focusing on contextual variables is even more important when we consider equity issues. Teachers are expected to deal with complex challenges in that children bring their social lives at home, on the street, in their communities, as well as their feelings about themselves as learners, to the classroom. Opportunities for teachers to learn becomes a critical issue at a time when increasing numbers of students are walking into classrooms who have been neglected, abused, or deprived, and whose experience with adults and schools set up and reinforce a sense of futility, which comes out of a long history of expectations of failure. Brewer, Stasz, and Mandel speak to the need to gauge excellence in teaching and to track the equitable distribution of opportunities to learn for teachers and students. The most daunting problem that we as educators face heading into the next century is one of achieving equity—students' equal access to knowledge and learning experiences. Gross inequities of resources have been documented in case studies by Kozol and in national data in the mathematics and science teaching fields by Oakes. More recently, my work with Darling-Hammond validates what we already know through anecdotal evidence—that is, students in poor and minority schools are taught by the least qualified and most inexperienced teachers.

Brewer and Stasz underscore the importance of ensuring equitable delivery of schooling for all students. Mandel calls attention to the underreported scandal in American education: in the face of shortages of qualified teachers, we are opting for short-term solutions by issuing emergency certificates, by lowering entry-level standards, and by misassigning teachers out of their fields. If disproportionate numbers of the least prepared elementary and secondary public school teachers teach in the most disadvantaged communities, it is likely that these children are experiencing fewer opportunities to learn than children in the most advantaged communities.

Thus, data that illustrate the interactions between teacher performance, student learning processes, and workplace supports may help us to deepen our understanding of the complexities of teaching and learning and also to document equitable delivery of schooling to all students.

References

- Darling-Hammond, L. and Sclan, E.M. 1996. "Who Teachers Are and What They Think." In T. Butterly and J. Sikula (eds.), *Handbook of Teacher Education*. Association of Teacher Educators.
- Kozol, J. 1992. *Savage Inequalities*. New York: Crown.
- Oakes, J. 1990. *Multiplying Inequalities: The Unequal Distribution of Mathematics and Science Opportunities*. Santa Monica, CA: Rand.
- Rosenholtz, S. 1991. *Teachers' Workplace*. New York: Teachers College Press.

MARY ROLLEFSON

David Mandel begins his paper with a “chicken or the egg” problem: Without a definition of the “elements of highly accomplished practice” we do not know how to collect data on teacher and teaching quality, but without data on teaching and student learning we can not define teacher quality. In other words without a definition, we can not collect data, and without data we can not validate a definition.

Responsibility for this situation is put on the teaching profession, but I believe that responsibility extends to the education research community as well, which has found little empirical support for the relationship between teacher and teaching variables and student outcomes. However, as Emerson Elliott mentioned at the conference, research from the University of Wisconsin Center on Authentic Pedagogy, which identifies qualities of teaching that relate to improvements in student performance, holds promise.

Mandel proposes the standards of the National Board for Professional Teaching Standards as the framework for better data on the quality of teachers and the institutions and the programs designed to provide them. In addition, he emphasizes the importance of data on teacher quality and suggests that all of our data on teachers are marginalized and diluted, without quality information that goes beyond our standard criteria of major or minor and certification in the teaching assignment field.

It is true, as Mandel says, that NCES “stubs its toe” on this issue and that quality data would vastly improve our understanding of teacher supply and demand—which ultimately is a quality issue, since imbalances are often resolved through adjustments in teacher quality. Another area where teacher quality data would improve understanding is in equity. That is, among different populations of students, which population has access to the type of quality teaching that makes a difference in learning? I suspect that important differences in the quality of teachers from different supply sources and in student access to good teaching are only hinted at in our current data on teachers.

Good measures of teacher quality would also help us understand how quality practice develops or fails to do so both in the course of teacher education and throughout a teaching career. The implications of this understanding for teacher education and continuing professional development are enormous.

Mandel also points out the importance of quality data for examining institutions of teacher education. In terms of a future direction for NCES, this may be important as well. I would suggest looking to NCATE standards and the work of INTASC (Interstate New Teacher Assessment and Support Consortium) for a framework. But if teacher quality is in its infancy, I suspect much work needs to be done in this area.

His recommendations to NCES are twofold: one is the use of the more traditional method of a statistical agency, the other relies on methods that are less traditional—case studies and assessments of teacher preparation programs. The latter calls into question what our appropriate role is, especially in a time of budget cuts. However, these approaches are more appropriately the territory of OERI and the wider education research community. In addition, Mandel points out the critical

connection between research and statistical data collection. I think NCES would eagerly await these results.

I believe that some of the more traditional recommendations hold promise. For instance, one is to collect teacher data every 5 years, which we already are doing in SASS, although the periodicity varies. Related to this is the recommendation to combine National Board information on board-certified teachers—rich information from teacher videos and commentaries on their teaching goals, context of instruction, student understanding, and the teacher rationale for the approach they use—with NCES data collection on these teachers, and to follow these teachers. Also, current plans have been made to include board-certified teachers in the next SASS, and perhaps they could also be followed in the Teacher Follow-up Survey. The combination of a sample of *quality* teachers in a data set that contains both rich National Board data and standard SASS data would provide us with an opportunity to do research on the issue of teacher quality. Finally, another good recommendation is to use board-certified teachers to help NCES assess its data on teacher quality.

Thus, as mentioned previously, without defining teacher quality we can not collect data and without data we can not validate the definition of teacher quality.¹ These two conditions define the ground for basic research; using that research as a framework, indicators of teacher quality can be developed to collect data from large nationally representative samples. This needs to happen before we can apply our usual methodologies to this resolving issue.

Mandel's paper raised several questions, which I would like to comment on. First, if National Board standards are a framework for defining teacher quality, then what are the dimensions of that framework, and what are some of its most important content areas? I would like to have had those laid out more explicitly in his paper. If we are not yet close to a working definition, then perhaps we can at least get a hint of where we are now and how far we have to go. I would also suggest adding some other places to search for definitions of teacher quality: for example, the NCATE standards, the work of INTASC (the Interstate New Teacher Assessment and Support Consortium), and some of the OTL research discussed in the Brewer and Stasz paper. Furthermore, we need more qualification about teacher quality—that is, we need data not just on the excellent, highly accomplished, and advanced end of the continuum, but on the whole distribution.

Finally, there were some more practical recommendations that were not discussed as thoroughly. These were to collect data on:

- Teachers' professional development—What they are doing to strengthen practice, and what kind of support are they receiving to do it?
- Public investment in teacher preparation and in continuing professional development—What kinds of investments and what kinds of programs? This points to the need for basic data on institutions that educate teachers.

Although these two items are provided almost as an afterthought, they point to important areas in which data could be improved in the near future. And they certainly deserve more discussion.

Notes

1. When developing a working definition of teacher quality, it would be useful to turn to the standards of the National Board for Professional Teaching Standards.

SHARON BOBBITT

On the first day of school in the fall, both new and experienced teachers go into their classrooms to meet their new students. They shut the door, take roll, and begin practicing what has been called “the second most private act”—teaching. While this stereotype is slowly changing, little is known about what happens inside the classroom walls, in the interaction of teacher and student, to bring about learning. The Brewer and Stasz paper could have been entitled “Illuminating the Black Box.” The paper makes a serious and credible attempt to help the National Center for Education Statistics figure out how to measure and report on what happens inside the black box of the elementary/secondary school classroom.

Brewer and Stasz use the terminology of Opportunity to Learn (OTL) to discuss data issues related to instructional practices and classroom processes. OTL started as a narrow concept in the context of international surveys. In reporting achievement scores in mathematics across many countries, a key variable was whether students in each country had been exposed to mathematical concepts in their curriculum by the time the assessment was conducted. Obviously, a student who has never had the opportunity to learn a mathematical concept will perform less well on those portions of the assessment that tap this concept than a student who has had extensive exposure to the concept in his or her school curriculum. The SIMS study asked teachers to rate whether or not their students had been exposed to the items in the assessment. This narrow definition of opportunity to learn was simply, therefore, a measure of students’ exposure to items on which they were being assessed.

Andy Porter and others have taken the concept of opportunity to learn and expanded it to encompass a much broader definition. Porter argues that a student’s opportunity to learn includes not only appropriate curriculum content but also high-quality pedagogy and adequate classroom resources. Students who have access to more of these three elements have, in some ways, more opportunity to learn challenging content. It is this framework of OTL that Brewer and Stasz adopt to discuss the measurement of classroom processes by NCES.

Using this OTL framework, Brewer and Stasz make four recommendations. First, they suggest that NCES enhance survey items on curriculum content and pedagogy. Second, they recommend enhancing our measurement of student learning through survey and collecting artifacts. Third, they suggest enriching classroom process data through alternative data collection methods. Finally, they suggest enhancing instructional resource measures. I would like to address each of these recommendations.

One of key issues about measuring curriculum content and pedagogy, especially in nationally representative surveys, is whether to make the questions subject- and grade-specific or try to formulate items that would be applicable to all teachers. Policy Study Associates (PSA), under contract to NCES, has recently struggled with this issue for the last administration of the Teacher

Follow-up Survey in school year 1994–95. This questionnaire is administered to a nationally representative sample of teachers of all subjects in grades K through 12. NCES wanted to include a module of questions on teacher instructional practices and classroom processes that would be applicable to the diverse group of teachers in the sample. Building on previous research in this area by Porter Burstein and McDonnell and others, PSA developed a set of questionnaire items that were tested in teacher focus groups before being administered in the TFS. It will be very interesting to see the amount of variation in a sample of over 7,000 teachers in basic instructional practices. The data will be available in the spring of 1996. While instructional practices may be able to be generalized to apply to all teachers, curriculum content must of necessity be subject- and grade-specific. NCES has also recently funded some additional work by Andy Porter on his teacher questionnaires about curriculum content for middle school math and science. By attempting to develop survey items that will be broadly applicable where possible and subject-specific when necessary, NCES hopes to improve the measurement of curriculum content and instructional practices in its large-scale surveys.

The second recommendation involves enhancing measurement of student learning through surveys and the collection of artifacts. As the authors correctly note, you cannot measure everything you want to know through large-scale surveys. Although it is very early in the process, NCES is experimenting with other forms of data collection to enhance our understanding of things that surveys cannot measure. The authors point to the work of Jim Stigler, who has videotaped teachers in their classrooms, as a promising possibility to build our capacity to measure what is going on inside the black box. It is also possible for NCES to fund, support, or conduct smaller scale research efforts that feed off of the large-scale sample surveys. Targeted subsamples (or even nationally representative subsamples) could be selected from existing sample surveys, such as the Schools and Staffing Survey, to investigate issues that are either too complex or too expensive (or both) to measure on the entire sample. Such subsamples could be candidates for using innovative data collection methodologies such as videotape (like Stigler), teacher logs, or classroom observation.

Thirdly, Brewer and Stasz recommend enriching classroom process data through alternative data collection methodologies. Working again with PSA, NCES has funded the development of survey items intended to measure instructional practices and classroom process in middle school mathematics. While the exact content of these items is not important, the process of development is. PSA is using teacher logs, classroom observation, and focus groups to validate the items that the teachers fill out in the survey form. The development of a validated module of items would enhance the ability of NCES to measure instructional practices in all of its large-scale surveys.

Finally, the paper recommends that we enhance our measures of instructional resources. I agree that this is one area where our surveys are weak, yet we have not made too much progress to date in working to fill this gap. One of the most outstanding teachers in the country recently told me at the Goals 1000 Teacher Forum that she gets \$40 each year from her school for classroom supplies. Teachers on the U.S. Department of Education listserv report routinely spending \$1,000 to \$2,000 of their own money each year for instructional supplies. Brewer and Stasz coin a wonderful term for this phenomenon, the “opportunity to teach.” NCES can add and should improve the measurement of the resources available for instruction in this country’s classrooms.

The Brewer and Stasz paper leaves us with several questions about the relationship of instructional practices and student outcomes. How do you validate measures of classroom processes and their impact on student learning, broadly speaking? Is it enough that teachers are doing the things in the classroom that the experts (for example, NCTM and NSTA) say they should be doing? Or does there need to be evidence that the pedagogy is resulting in improved student outcomes? Are instructional practices, in and of themselves, worth studying? Hopefully, as we improve our ability to measure classroom processes, we will be able to understand better the complex relationship of curriculum content, instructional pedagogy, and resources, and we will be able to get a better glimpse inside the black box of the classroom.