7 New Data Collection Methodologies, Part I: Observational Strategies

THIS PAGE INTENTIONALLY LEFT BLANK

Large-Scale Video Surveys for the Study of Classroom Processes

James W. Stigler

INTRODUCTION

In thinking about what kinds of indicators NCES might employ in the next 10 years it is useful to consider the kinds of information that might be important to improve education. NCES might collect three broad classes of information: 1) data on outcomes, whether related to achievement, attainment, or other goals; 2) data on policy implementation, i.e., data that indicate whether or not educational policies have been implemented, and where implemented how effective the policies are; and 3) data relevant to the processes that produce educational outcomes.

All three types of data are important for the improvement of student learning and achievement. However, it is my view that too much emphasis has been placed on the measurement of outcomes, and not enough on the study of processes that cause the outcomes. The critique that W. Edwards Deming leveled at American industry applies just as well to American education: quality cannot be improved simply by mass inspection of products. Instead, it is necessary to reflect on the processes that produce quality products, and then take measures to bring those processes under control. Likewise in education, we cannot improve student learning simply by measuring outcomes; we must investigate the processes that lead to high student achievement.

Chief among the processes that cause student achievement must surely be the processes of teaching and learning that transpire inside classrooms. Yet, surprisingly, we collect virtually no data—whether at the national, state, or local levels—that yield information about what is going on in classrooms. This is not because such data are deemed unimportant: in a series of papers commissioned by the National Center for Education Statistics (NCES) in 1985, papers designed to set the agency's priorities for the next 10 years, the need for classroom process indicators was raised numerous times (Hall, Jaeger, Kearney, and Wiley 1985). Cronin (1985), for example, expressed concern with the paucity of data that could document curricular breadth or the actual implementation of curricular reform in the classroom. Moreover, Peterson (1985) cited a near complete lack of data on the quality of educational activities in the nation's classrooms, or even on the time teachers devote to various instructional activities. Including such indicators in the future was a clear recommendation of the 1985 report.

Ten years later, such indicators are still deemed important, but they are still lacking. A new NCES survey of leading educators and researchers, conducted by MPR Associates in the summer of 1994, again finds that the most frequently cited area in which better national data are needed is that

of instructional practice. Yet the NCES *Condition of Education 1994* shows virtually no information at all concerning what happens in classrooms.

Probably the main reason for the continued lack of classroom process indicators is that what happens in classrooms is very difficult to describe and measure, especially on a large scale. What measures we do have are largely based on questionnaires in which teachers report on what happens in their own classrooms. Yet using questionnaires to measure classroom processes is problematic, as will be discussed below. Observation, on the other hand, would seem the natural way to study classroom processes. But observation is notoriously difficult and labor intensive.

Overview of This Paper

The first section of this paper will present a plea for the development of observational indicators of classroom process. The discussion will focus on what can be learned from observation, and argue for the advantages of video over live observers. The next section will explain some of the methodological issues that arise when video is used on a large scale. The final section of the paper will discuss the TIMSS Videotape Classroom Study, which I believe is the first attempt to use video for studying nationally representative samples of classroom teachers. This description will be detailed because the study really is the first of its kind, and much of what we have learned in this study will be helpful to those who follow. The software system we have developed for use on the project will also be described here.

Most researchers, on hearing the word "video," imagine a small-scale qualitative study. What I hope to demonstrate is the promise of using video for large-scale studies in which qualitative information can be easily combined with quantitative indicators.

WHAT WE CAN LEARN FROM CLASSROOM OBSERVATIONS

Having decided to study the processes of teaching and learning that go on inside classrooms, we must next decide how best to study these processes. In this section, a case will be made for using classroom observations, first by outlining the disadvantages of traditional questionnaire measures, and then by discussing the kinds of information that can be collected in observational studies. The focus here will be on two broad goals we might have for observational studies: first, to develop empirically validated models of instructional quality together with indicators for assessing instructional quality; and second, to monitor the implementation and effectiveness of educational policies.

Limitations of Questionnaires for Studying Classroom Processes

Most attempts to measure classroom processes on a large scale have used teacher questionnaires. Teachers have been asked, for example, to report on the percentage of time they spend in lecture versus discussion, the degree to which problem solving is a focus in their mathematics classrooms, and so on.

There are at least three major limitations imposed by the use of questionnaires to study classroom instruction. First, the words researchers use to describe the complexities of classroom instruction may not be used in the same way by teachers, or in a consistent way among different teachers. The phrase “problem solving” is a good example. Many reformers of mathematics education call for problem solving to become the focus of the lesson. But different teachers interpret this phrase in different ways. For instance, one teacher may believe that working on word problems is synonymous with problem solving, even if the problems are so simple that students can solve one in 15 seconds. Another teacher may believe that a problem that can be solved in less than a full class period is not a real problem but only an exercise. This kind of inconsistency is the rule in this country, where teachers have few opportunities to observe or be observed by other teachers in the classroom. Because teacher training in the United States generally does not engage teachers in discussions of classroom instruction, and because they are often isolated from one another by the conditions under which they work, teachers do not develop shared referents for the words used to describe instruction. Thus, although teachers may fill in questionnaires about their teaching practices, interpreting their responses is problematic.

A second problem with relying on questionnaire-based indicators of instruction concerns their accuracy. Even if teachers do interpret a question consistently, they may be inaccurate in reporting on processes that are probably at least in part outside of their awareness. Teaching is part planning, part performance. Teachers may be accurate reporters of what they planned for a lesson (e.g., what kind of demonstration they used to introduce the lesson), but they may be inaccurate when asked to report on actual aspects of teaching. Teachers process enormous quantities of information during a typical lesson and must continually adapt to changing circumstances, a process that happens too quickly to be under the teacher's conscious control. Observational studies of gender bias in teachers' questioning generally surprise teachers with their results: teachers who call on boys more frequently than girls, for example, have no idea that this is happening. Obviously, they would not be able to identify such a bias on a questionnaire.

A third limitation of questionnaires is their static nature. Teachers can only answer the questions we as researchers were clever enough to ask. Where an observer might notice something significant just by being in the classroom, questionnaires could not lead to the generation of new ideas or hypotheses in the same way.

Developing and Assessing Models of Instructional Quality

Developing observational indicators of classroom processes could serve two primary purposes: first, to aid in developing models of instructional quality; and second, to monitor and evaluate the effectiveness of educational policies.

Classroom instruction is a complex and multidimensional process. Nevertheless, we must have theoretical and methodological tools for studying classroom instruction if we are to improve it. Observational studies make it possible to develop indicators of classroom instruction that can then be used to develop and validate models of instructional quality. If this effort is to succeed, a number of indicators must be combined: we must examine the content of classroom lessons (the so-called implemented curriculum) as well as the methods teachers use to engage students in the content. That is, we must be able to examine the planned/structural aspects of instruction as well as the on-line

implementation of instruction that occurs as the lesson unfolds. Evolving models of instructional quality will be linked to improved indicators for assessing instructional quality.

Monitoring and Evaluating Educational Policies

Once consensus emerges on classroom-based definitions of quality instruction, policies designed to improve the quality of instruction will emerge based on these definitions. Another role of observational studies, therefore, will be to monitor the implementation of these policies in classrooms, and to assess their effectiveness.

Policies designed to improve instructional quality will be similar to opportunity-to-learn (OTL) standards. As described by Porter (1995), these standards will offer two distinct advantages over outcome-based standards alone: 1) they can provide a vision of what good practice looks like; and 2) they can provide a system of school process indicators related to OTL goals.

A good example of these new policies is contained in the NCTM Standards, which represent a consensus on what high-quality instruction should look like in the classroom. Operationalizing this consensus in a system of classroom-based observational indicators will allow us to assess the degree to which the standards are being implemented, and to empirically assess the effectiveness of the teaching practices described in them.

ADVANTAGES OF VIDEO OVER LIVE OBSERVATION

Video has distinct advantages over live observation in the study of classroom processes. The next section will present these advantages.

Enables Study of Complex Processes

Classrooms are complex environments, and instruction is a complex process. Live observers are necessarily limited in what they can observe, and this, in turn, limits the kinds of assessments they can do. With video, the problem of "bandwidth" becomes manageable: observers can code video in multiple passes, coding different dimensions of classroom process on each pass. On one pass, for example, they might code the ways materials are used, on another the behavior of students.

Not only can coding be done in passes but it also can be done in slow motion. With video, for example, it is possible to transcribe the language of the classroom, enabling far more sophisticated analysis of complex discourse processes. Detailed coding of classroom discourse would be unthinkable without the capacity to slow down and listen again.

Increases Inter-Rater Reliability, Decreases Training Problems

Video also resolves problems of inter-rater reliability that are difficult to resolve in the context of live observations. Although it is possible to send observers out in pairs for the purpose of assessing reliability of indicators, it is often very inconvenient to do so. For example, if a study is being performed cross-culturally, or in geographically distant locations, it is often necessary to hire local observers. Bringing these observers together to check reliability is not usually feasible.

Having video also makes it far easier to train observers. With video, inter-rater reliability can be assessed not only between pairs of observers but also between all observers and an expert "standard" observer. Disagreements can be resolved based on re-viewing the video, making such disagreements into a valuable training opportunity. And, the same segments of video can be used for training all observers, increasing the chances that coders will use categories in comparable ways.

Amenable to Post-Hoc Coding, Secondary Analysis

Most survey data sets lose their interest over time. Researchers decide what questions to ask, and how to categorize responses, based on theories that are prevalent at a given time. Video data, because they are "pre-quantitative," can be re-coded and analyzed as theories change over time, giving these data a longer shelf life than other kinds. Researchers in the future may code videotapes of today for purposes completely different than those for which the tapes were originally collected.

Amenable to Coding from Multiple Perspectives

For similar reasons, video data are especially suited for coding from multiple disciplinary perspectives. Tapes of mathematics classes in different countries, for example, might be independently coded by psychologists, anthropologists, mathematicians, and educators. Not only is this cost effective but also it facilitates valuable communication across disciplines. The most fruitful interdisciplinary discussions result when researchers from diverse backgrounds compare analyses based on a common, concrete referent.

Merge Qualitative and Quantitative Information

Video makes it possible to merge qualitative and quantitative analyses in a way not possible with other kinds of data. With live-observer coding schemes the qualitative and quantitative analyses are done sequentially: initial qualitative analyses lead to the construction of the coding scheme, and implementation of the coding scheme leads to a re-evaluation of the qualitative analysis.

When video is available, it is possible to move much more quickly between the two modes of analysis. Once a code is applied, the researcher can go back and look more closely at the video segments that have been categorized together. This kind of focused observation makes it possible to

see, for example, that the segments differ from each other in some significant way, and this difference may form the basis for a new code.

It also is possible with video to use example segments in reporting the results of the research. This gives the consumers of the information a richer qualitative sense of what each category in the coding system means.

Video Provides Referents for Teachers' Descriptions

Mentioned earlier was the problem that teachers lack a set of shared referents for the words they use to describe classroom instruction. Video, in the long run, can provide teachers, as potential consumers of the research, with a set of such referents. Definitions of instructional quality and the indicators developed to assess instructional quality can be linked to a library of video examples that teachers can use in the course of their professional development. In the long run, a shared set of referents can lead to the development of more efficient and valid questionnaire-based indicators of instructional quality.

A Source of New Ideas

A final advantage of video over other kinds of data is that it becomes a source of new ideas on how to teach. Because these new ideas are concrete and grounded in practice, they are potentially immediately useful for teachers. Questionnaires and coding schemes can help us to spot trends and relationships, but they cannot uncover a new way of teaching the Pythagorean Theorem. Video, especially if collected on a large scale, can be a treasure chest of such ideas.

ISSUES IN VIDEO RESEARCH

The next section will cover a number of issues that must be resolved in order to conduct meaningful video research.

Standardization of Camera Procedures

Left to their own devices, different videographers will photograph the same classroom lesson in different ways. One may focus in on individual students, while another may shoot wide shots in order to give the broadest possible picture of what is happening in the classroom. Yet another might focus on the teacher or on the blackboard. Because the intention is to study classroom instruction, not the videographers' camera habits, it is important to develop standardized procedures for using the camera, and then to carefully train videographers to follow these procedures.

The Problem of Observer Effects

Given that the camera is used in a consistent way, we must next consider the possible effect the camera might have on what happens in the classroom. Will students and teachers behave in

typical fashion with the camera present, or will we get a view that is biased in some way? Might a teacher, knowing that she or he is to be videotaped, even prepare a special lesson just for the occasion that is unrepresentative of normal practices?

This problem is not unique to video studies. Questionnaires have the same potential for bias: teachers' questionnaire responses, as well as their behavior, may be biased toward cultural norms. On the other hand, it may actually be easier to gauge the degree of bias in video studies than in questionnaire studies. Teachers who try to alter their behavior for the videotaping will likely show some evidence that this is the case. Students, for example, may look puzzled or may not be able to follow routines that are clearly new for them.

It also should be noted that changing the way a teacher teaches is notoriously difficult to do, as much of the literature on teacher development suggests. It is highly unlikely that teaching could be improved significantly simply by placing a camera in the room. On the other hand, teachers will obviously try to do an especially good job, and may even do some extra preparation, for a lesson that is to be videotaped. We may, therefore, see a somewhat idealized version of what the teacher normally does in the classroom.

Minimizing Bias Due to Observer Effects

We have identified three techniques for minimizing bias due to videotaping. First, instructions must be standardized. Teachers generally do not want to bias the results of a study, but may inadvertently do so in an effort to help researchers. It is important, therefore, to clearly communicate the goal of the research to the teacher in carefully written, standard instructions. The teacher, when properly informed, becomes an important ally in the effort to get unbiased results. Teachers need to be told that the goal is to videotape a typical lesson, whatever they would have been doing had the videographer not shown up. Teachers can also be explicitly asked to prepare for the target lesson just as they would for a typical lesson.

A second technique is to assess the degree to which bias has occurred. After the videotaping, teachers can be asked to fill out a questionnaire in which they rate, for example, the typicality of what we see on the videotape, and describe in writing any aspect of the lesson they feel was not typical. We also can ask teachers whether the lesson in the videotape was a stand-alone lesson or part of a sequence of lessons, and to describe what they did yesterday and what they plan to do in tomorrow's lesson. Lessons that are stand-alone and that have little relation to the lessons on adjoining days may be special lessons constructed for the purpose of the videotaping. In the work we have done, however, this is rarely the case.

Finally, we must use common sense in deciding the kinds of indicators that may be susceptible to bias, and take this into account in interpreting the results of a study. It seems likely, for example, that students will try to be on their best behavior with a videographer present, and so we may not get a valid measure from video of the frequency with which teachers must discipline students. On the other hand, it seems unlikely that teachers will ask different kinds of questions while being videotaped than they would ask when the camera is not present.

Sampling and Validity

Observer effects are not the only threat to validity of video survey data. Sampling—of schools, teachers, class periods, lesson topics, and parts of the school year—is also a major concern.

One key issue is the number of times any given teacher in the sample should be videotaped. This obviously will depend on the level of analysis to be used. If we need a valid and reliable picture of individual teachers then we must tape the teacher multiple times, as teachers vary from day to day in the kind of lesson they teach as well as in their success in implementing the lesson. If we want a school-level picture, or a national-level picture, then we obviously can tape each teacher fewer times, provided we resist the temptation to view the resulting data as indicating anything reliable about the individual teacher.

On the other hand, taping each teacher once limits the kinds of generalizations we can make about instruction. Teaching involves more than constructing and implementing lessons. It also involves weaving together multiple lessons into units that stretch out over days and weeks. If multiple teachers are taped once, it will be difficult to code the dynamics of teaching over the course of a unit. Inferences about these dynamics cannot necessarily be made, even at the aggregate level, based on one-time observations.

Another sampling issue concerns representativeness of the sample across the school year. This is especially important in cross-national surveys where centralized curricula can lead to high correlations of particular topics with particular parts of the year. Although at first it may seem desirable to sample particular topics in the curriculum in order to make comparisons more valid, in practice this is virtually impossible. Especially across cultures, teachers may define topics so differently that the resulting samples become less rather than more comparable. Randomization appears to be the most practical approach to ensuring the comparability of samples.

Confidentiality

Unlike traditional data sets, much of the contents of video data will still be unanalyzed by the time a public-use data set is constructed. Yet, the fact that images of teachers and students appear on the tapes makes it even more difficult than usual to protect the confidentiality of study participants. An important issue, therefore, concerns how procedures can be established to allow continued access to video data by researchers interested in secondary analysis.

One option is to disguise the participants by blurring their faces on the video. This can be accomplished with modern-day digital video editing tools, but it is expensive at present to do this for an entire data set. A more practical approach is to define special access procedures that will make is possible protect the confidentiality of participants while still making the videos available as part of a restricted-use data set. (One such set of procedures is outlined below.)

Expense/Logistics

Video surveys can be far more expensive than traditional surveys. In fact, the future viability of such studies will depend on our ability to manage the considerable expense and logistical challenges posed by such studies.

Contrary to traditional surveys, which require intensive and thorough preparation up front, the most expensive and daunting part of video surveys is in the data management and analysis phase. Whereas information entered on questionnaires can easily be transformed into computer readable format, such is not the case for video images. Thus, it is necessary to find a means to index the contents of the hundreds of hours of tape that can be collected in a video survey. Otherwise, the labor involved in analyzing the tapes grows enormously.

Once data are indexed, there is still the problem of coding. Coding of videotapes is renowned as highly labor intensive. But there are strategies available for bringing the task under control. One approach to this task will be elaborated below.

TIMSS VIDEOTAPE CLASSROOM STUDY: SCALING UP TO VIDEO SURVEYS

Having discussed both the opportunities and the challenges offered by video surveys, we now turn to briefly describe an example of such a survey that is currently underway. This study, which is part of the Third International Mathematics and Science Study (TIMSS), represents an unprecedented attempt to use video in a national-level survey research context. Focused on 8th-grade mathematics, the study compares the teaching practices of German, Japanese, and American teachers. Data collection is complete; we are now coding the data. All of the issues described above have been encountered in the conduct of this study. Our experiences in addressing these issues will hopefully be instructive as we contemplate future video surveys.

Introduction to the Study

Background and Objectives

TIMSS is the third in a series of international studies conducted under the auspices of the International Association for the Evaluation of Educational Achievement. The first two of these studies (Husen 1967; McKnight et al. 1987) established large cross-national differences in achievement, and provided some information on contextual factors, such as curriculum, that could be related to the achievement differences.

Perhaps because students from the United States did relatively poorly in the first two studies, the U.S. sponsors of TIMSS (primarily NCES) have placed a high priority on improving the quantity and quality of contextual information to be collected in TIMSS. Predicting that the performance of U.S. students would continue to be low relative to other industrialized countries, the U.S. Department of Education has tried to ensure that the results of TIMSS bear not only on the achievement of students but also on the processes that lead to achievement. The goal is to make TIMSS more useful to policymakers than either of the first two IEA studies have been.

In accordance with this goal, NCES has funded two studies to complement the main TIMSS data. Both of these studies focus on three countries: Germany, Japan, and the United States. The first involves comparative case studies of various aspects of the educational systems of each country. The second is the Videotape Classroom Study.

The goal of the Videotape Classroom Study is to provide a rich source of information on how 8th-grade mathematics is taught in Germany, Japan, and the United States. This is the first large-scale study to collect videotaped records of classroom instruction in the mathematics classrooms of different countries. The study has four main objectives:

1) To develop objective observational measures of classroom instruction that will serve as valid quantitative indicators, at a national level, of teaching practices in three countries;

2) To complement information about classroom instructional methods collected by the TIMSS background questionnaires with information gained from actual classroom observations in order to obtain a richer description of classroom teaching practices in Japan, Germany, and the United States;

3) To compare actual mathematics teaching methods in the United States and other countries with those recommended in current reform documents and with teachers' perceptions of those recommendations; and

4) To assess the feasibility of applying videotape methodology in future wider scale national and international surveys of classroom instructional practices.

Design of the Study

National probability samples of 8th-grade mathematics classes from Germany, Japan, and the United States are participating in the study. The samples are random subsamples of the TIMSS main study sample, which is selected according to the TIMSS sampling plan. The plan was to sample 100 classrooms from Germany and the United States, and 50 from Japan. The final sample consists of 100 classrooms from Germany, 81 from the United States, and 50 from Japan.

The video study includes two major sources of data: videotapes and questionnaires. In addition, supplementary materials helpful in understanding the lesson, such as examples of textbook pages and worksheets, were collected. Each classroom was videotaped once on a date convenient for the teacher. One complete lesson—as defined by the teacher—was videotaped in each classroom. One videographer was employed in each country. In Germany and the United States videotaping was carried out over a 7-month period, and in Japan, over a 4-month period. Teachers were told that we wanted to tape a “typical” lesson and, thus, that they should do no special preparation on the day of taping. After the taping, each teacher was given a questionnaire and an envelope in which to return it. The purpose of the questionnaire was to assess how typical the lesson was according to the teacher, and to gather contextual information important for understanding the contents of the videotape. Both taping procedures and questionnaire contents are described in more detail below.

The LAVA Software System

To facilitate the processing of such large quantities of video data, we decided to digitize all of the video and supplementary materials, which allowed them to be stored, accessed, and analyzed by computer. Each lesson videotape was digitized, compressed, and stored on CD-ROM disks, one lesson per disk. We then designed and built a multimedia database software application that would enable us to organize, transcribe, code, and analyze the digital video. This interactive video analysis system, which we have called LAVA (for LA Video Analysis), represents a major advance in technology available to aid in the implementation of video surveys. For this reason, the system will be described in some detail along with the description of each part of the study.

Digital video offers several advantages over videotape for use in video surveys. First, the resulting files are far more durable and long-lasting than videotape. CD-ROM disks are assumed to last for 100 years, as opposed to a much shorter lifespan for videotape. Digital video files also can be copied without any loss in quality, which again is not true for videotapes. And, digital files will not wear out or degrade with repeated playing and replaying of parts of the video. Digital video also enables random, instantaneous access to any location on the video, a feature that makes possible far more sophisticated analyses than are possible with videotape. For example, when coding a category of behavior, it is possible to quickly review the actual video segments that have been marked for that category. This rapid retrieval and viewing of coded segments makes it possible to notice inconsistencies in coding, or to discover new patterns of behavior, that would not be possible without such access.

The LAVA software system consists of several modes. Transcribe mode is used for transcribing the videotapes. Code mode allows users to define categories and code them across a large number of videos. Analyze mode is used to search the database and retrieve video segments on the basis of transcript or codes, and to produce spreadsheet outputs of data that can be imported into standard statistical analysis programs. These modes will be described in more detail later.

Instructions and Questionnaire

As pointed out earlier, both instructions to the teacher and the questionnaire that accompanies the videotaping are means of minimizing the potential bias of observer effects. Designing each of these was given careful consideration in the TIMSS video study.

Instructions

It is not feasible to show up unannounced to videotape classroom lessons. Because teachers know when the taping is to take place, they undoubtedly prepare for it in some way. How they prepare probably will have an impact on the kind of instruction we see. Teachers may try to teach like they think we want them to teach; they almost certainly will try to do what they believe is a good job.

In order to cut down somewhat on the variability in preparation methods across teachers, we gave teachers in each country a common set of instructions for how we wanted them to prepare. Teachers were told the following:

> Our goal is to see what typically happens in American mathematics classrooms, so we really want to see exactly what you would have done had we not been videotaping. Although you will be contacted ahead of time, and you will know the exact date and time that your classroom will be videotaped, we ask that you *not* make any special preparations for this class. So please, do not make special materials, or plan special lessons, that would not typify what normally occurs in your classroom. Also, please do not prepare your students in any special way for this class. Do not, for example, practice the lesson ahead of time with your students.

Questionnaire

The purpose of the teacher questionnaire was to elicit information that would help us in the analysis and interpretation of the videotapes. Items for the questionnaire were generated by project personnel in consultation with persons working on the main TIMSS questionnaire, questionnaire design specialists from Westat, mathematics educators, and classroom teachers. Questions were edited and selected to yield a questionnaire that would take approximately 20–30 minutes for teachers to complete.

The questionnaire was translated into German and Japanese, translated back into English, and then pilot-tested on teachers participating in the field test. The responses from the field test were discussed by German, Japanese, and American collaborators, and based on these discussions the questionnaire was revised.

The final translation of the questionnaire was painstakingly reviewed, question by question, by a group of German, Japanese, and American researchers, each of whom was fluent in two of the three languages. Questions that were judged too difficult to translate accurately were dropped from the questionnaire.

The resulting questionnaire consists of 3 parts with a total of 28 questions. In Part A, we ask questions about the lesson that was videotaped, and about how the class was constituted and who the students were. In Part B, we ask the teachers to compare what happened in the videotaped lesson with what would typically transpire in their classroom. In Part C, we ask teachers to describe what they know about current ideas on mathematics teaching and learning, and ask them to evaluate their own teaching in the videotape in light of these current ideas.

The information collected in the questionnaire will serve three purposes. First, information from the questionnaire will help us assess the quality and comparability of our samples across the three countries. Although teachers will be instructed not to prepare in any special way for the videotaping, we cannot take it for granted that what we see on the videotape is typical of what normally happens in a given classroom. Teachers thus will be asked to directly rate the typicality of the videotaped lesson, and these ratings will be compared across countries. Similarly, we will assess

the comparability of the samples across the three countries along several important dimensions. For example, whether a lesson deals with new material or review might be expected to influence the kind of teaching technique used. Knowing the percentage of lessons in each country that are new versus review will help us to judge the comparability of the samples.

A second purpose for the questionnaire is to provide coders with information that will help them interpret what they see on the videotapes. For example, it is often necessary to know the teacher's goal for a lesson in order to make sense of the activities that constitute the lesson, and so we ask the teacher to say what her or his goal for the lesson was. Similarly, to interpret the meaning a specific question has for students it is often helpful to know whether the question probes new material or reviews previously learned information. Again, teachers are asked to categorize the content of the lesson in this way on the questionnaire.

Third, the questionnaire responses will, in some cases, enter directly into the analyses—statistical and qualitative—of the videotapes. This will occur in several ways. First, questionnaire responses will enter into correlational analyses within each country to help us relate contextual factors to variations in classroom instruction. For example, we can investigate the degree to which instructional techniques vary according to the ability level of students in the class. Second, we can use questionnaire responses to identify sampling biases that may affect our results. For example, if lessons that deal with new material (as opposed to review material) are sampled more in one country than another, this information could be used as a covariate to correct for the bias in sampling. Third, by asking teachers to comment on the lesson that was videotaped, we can learn more about how teachers interpret the language of reform in mathematics education. For example, if a teacher tells us that his or her lesson was focused on problem solving, we can look at the video to see what the teacher meant by the term “problem solving.”

Filming in Classrooms

Before we could collect our first videotape, we had to accomplish a number of tasks. We had to 1) develop procedures for videotaping in classrooms that could be applied in comparable ways across three different cultures; 2) develop and implement methods for training videographers to use these procedures in a consistent way; and 3) evaluate the success of our training by comparing camera use across our three videographers. The following will describe how we accomplished each task.

Establishing Comparable Procedures

The success of any video survey will hinge on the quality and comparability of the tapes collected. What we see on video is not only dependent on what transpires in the classroom but also on the way the camera is used. If our aim is to compare certain aspects of instruction, then we must make sure that these aspects are clearly captured on all the tapes. In addition, we want to make sure that we are comparing classroom instruction, not camera habits. There are many decisions that must be made by the camera operator; if these are not made in a standardized manner, then the resulting tapes will not be comparable across classrooms or countries.

We developed procedures for camera use in collaboration with Scott Rankin, an experienced videographer who had worked with us on previous projects and who was therefore familiar with the challenges of documenting classroom instruction. Our goal was to develop a set of general principles and rules of thumb that would be easy for our videographers to learn, yet comprehensive enough to apply in any classroom situation. Of course, there are many rules and principles one could come up with depending on the goals of any particular survey. Reviewing ours, however, will at least serve to highlight the kinds of issues that must be considered when developing procedures for camera use. They might also be applicable to other studies.[1]

One camera was used, which of course limits the amount of information that can be collected. This constraint was imposed by NCES as a cost-saving measure, though it also makes the process of coding and analysis simpler than it would be with two cameras. The procedures for camera use presented would need to be altered if two cameras were used.

Basic Principles for Documenting Classroom Lessons

Because we wanted to see each lesson in its entirety, all videotaping was done in real time: the camera was turned on at the beginning of the class, and not turned off until the lesson was over. This means that we can study the durations of classroom activities by measuring their length on the videotape. Obviously, this would not be possible if there were any gaps in the recording.

Classrooms are complex environments where much is going on at any given time; it is impossible to document everything, particularly when only one camera is used. We decided on two principles to guide videographers in their choices of where to point the camera. These principles yield a comprehensive view of the lesson being taped.

Principle #1: Document the perspective of an ideal student. Assume the perspective of an ideal student in the class, then point the camera toward that which should be the focus of the ideal student at any given time. An ideal student is one who is always attentive to the lesson at hand, and always occupied with the learning tasks assigned by the teacher. An ideal student will attend to individual work when assigned to work alone, will attend to the teacher when he or she addresses the class, and will attend to peers when they ask questions or present their work or ideas to the whole class. In other words, we chose to point the camera so as to capture the experience of a student who is paying attention to the lesson as it unfolds. In cases where different students in the same class are engaged in different activities, the ideal student is assumed to be doing whatever the majority of students are doing.

Principle #2: Document the teacher. Regardless of what the ideal student is doing, be certain to capture everything that the teacher is doing to instruct the class. Usually the two principles are in agreement: whenever the ideal student is attending to the teacher, both principles would involve having the camera pointed at the teacher. However there are times when the two principles are in conflict. Take, for example, a case where the majority of students are doing seatwork while the teacher is working privately with two students at the board. The ideal student would be focused

on his or her work, not on the teacher. In situations like this one, the videographer must go beyond these two basic principles in order to determine where to point the camera.

The Exceptions: Three Difficult Situations

We have identified three common situations where the principles alone cannot guide choices about what to capture on the videotape. These situations are 1) when the ideal student would be focused on something other than the teacher, 2) when two speakers who are having a conversation will not fit in a single shot, and 3) when a speaker and an object being discussed will not fit in a single shot. We have developed a set of guidelines so that videographers will chose similar (i.e., comparable) shots when faced with each of these situations, and so that these shots will contain a maximum amount of useful information. The rest of this section presents a more detailed discussion of these situations and how to film them.

Situation #1: When the ideal student is not watching the teacher. As already mentioned, there are times when the ideal student should be attending to something other than the teacher. This most often occurs when students are given a task to work on individually or in small groups. Teachers can use this time in different ways. Sometimes they will walk around the class and monitor students' work. This is ideal from the videographers' point of view because by following the teacher with the camera one can also get a sense of what students are doing. In some instances, however, a problem arises because the teacher does not circulate around the class, but rather stays at the board or his or her desk. In such cases, the camera would need to be pointed in two different directions (toward the teacher and toward the students) in order to capture both the teacher and the focus of the ideal student.

Videographers were instructed to handle such situations by alternating between these two points of view. They were told to slowly do a sweep of the classroom by panning away from the teacher and then panning back to the teacher so as to document what the students are doing. After this sweep, they were told to focus on the teacher unless the nature of the students' activity changes in any significant way (e.g., new materials are introduced or they break into groups). If the students' activity were to change, videographers were instructed to carry out another sweep of the students, and then return to the teacher.

Situation #2: When two speakers will not fit in a single shot. A second difficult situation occurs when the teacher is conversing with a student (or a student is conversing with another student) and the two speakers are far enough apart so that they do not fit in a single camera shot. This often occurs when a teacher calls on a student seated in the back of the room, and then proceeds to converse with the student.

In this case, videographers were instructed to move the shot from speaker to speaker as they take turns talking. An exception to this rule occurs when one of the speaker's turns is so brief that there is no time to shift the camera before the turn is over. In this case, the camera should be kept on the person doing the most talking.

Situation #3: When the speaker and the object being discussed will not fit in a single shot. Another difficult situation occurs when a speaker and an object he or she is discussing will not

both fit into a single camera shot. This happens frequently, for example, when someone is talking about things written on the chalkboard or about concrete representations of a mathematical situation or concept.

In this kind of situation, videographers were told to document the object for long enough to provide the visual information needed to make sense of the talk, then to keep the shot on the speaker. For example, if the teacher is talking about a problem on the blackboard, the videographer should first tape the problem, then move to the teacher.

There is one important exception to this rule. Sometimes it is not sufficient to briefly see the object and then move to the speaker because the talk will make no sense unless one is seeing the object as it is being talked about. For example, if the speaker is pointing to specific features of the object as he or she talks, and if the direction of the points must be seen in order to understand the talk, then the rule is that the camera must stay on the object so that the talk can be understood.

How Close to Frame the Shot

Aside from making sure that videographers point their cameras at comparable things, we also wanted to make sure that their shots are framed in comparable ways. An extreme close-up of the teacher talking would provide a very different sense of the action taking place than a wide shot where the teacher is seen in the context of the classroom.

We decided that in general we wanted the widest shot possible, a shot professional videographers call the "Master of Scene" (MOS) or, more simply, the "master shot." From an aesthetic point of view closer shots often look better. However, the MOS provides more contextual information and thus was judged more appropriate for our purposes. The master shot also is less prone to bias because it does not artificially focus the viewer in on whatever aspect of the lesson the videographer judged to be most interesting.

Sometimes, however, there is crucial information that cannot be captured in a master shot. Common examples include objects being discussed during the lesson, or things written on the blackboard. In such instances, the camera should zoom in close enough to capture this information. In other words, although our preferred view of the classroom is the MOS, a closer shot must be used when it is needed to understand what is going on. Videographers were told to hold close shots long enough to enable a viewer to read or form a mental image of the information.

Moving from Shot to Shot

Finally, having devised guidelines for what to include in the shot, we also needed some rules for how to move from shot to shot. This, too, must be done in a standardized way if the tapes are to be fully comparable.

The guidelines we gave to the videographers were based on principles of good camera work. We taught them how to compose shots and execute camera movements in ways that follow basic

cinematographic conventions and fundamentals of good composition. Aside from wanting them to follow the same conventions, we wanted them to carry out good camera work. Bad camera work calls attention to itself and distracts the viewer from the contents of the tape.

Training Videographers

In order to make sure that the rules were applied correctly and reliably, we had to work intensively with the videographers. Each videographer participated in two training sessions, both of which were conducted by our professional videographer. The first training session lasted 9 days in the spring of 1994, after which each videographer was sent out to collect ten practice tapes for a field test. The second training session lasted 5 days and was held in the early fall of 1994. Following this second training session, videographers were given a test, and then sent off to collect the data.

We designed the training sessions with two goals in mind: First, we wanted to teach the videographers our camera use rules to the point that they could follow them second nature. In an actual taping situation, videographers would have to make rapid decisions about where to point the camera without time for reflection. Second, we wanted the videographers to learn and practice the fundamental skills of camera use. These skills include, for example, changing from one camera angle to another quickly without losing a focused image, tracking moving objects without having the object leave the shot, and moving rapidly back and forth from close-ups to master shots, while ending up centered on the shot that needs to be captured.

The first training session was devoted to five activities: Learning to use the equipment, practicing basic principles of good camera work, presentation and discussion of the standardized rules for taping classrooms, practice taping in mock classrooms, and practice taping in real classrooms. Activities in the second training session included reviewing and discussing the rules, critiquing practice tapes, and more practice taping in mock classrooms. A monitor hooked to the camera during the training sessions allowed videographers to rotate between practicing with the camera and watching/critiquing their peers in collaboration with the instructor.

The following is a helpful hint for others contemplating this kind of work. One has two alternatives in deciding who to hire and train as a video survey videographer: one can hire scientists (i.e., educational researchers) and train them to take good pictures, or one can hire artists (i.e., photographers) and teach them the importance of following standardized rules for camera use. In my experience, the latter is far easier, and the pictures are much more aesthetically pleasing.

Evaluating the Comparability of Camera Use

At the end of the second training session, we gave each videographer a test to measure and document how well they had internalized all they had been taught. A 7-minute mock lesson was created that covered many of the situations videographers needed to know how to handle. The lesson was taught three times, each one identical to the others, and was taped each time by one of the three videographers. The resulting tapes were analyzed and evaluated to make sure that our videographers would shoot lessons in a standardized manner.

To evaluate the videographers' performance on the test, we first produced a description of how the test lesson should have been videotaped. We listed the 22 events that took place in the lesson, and then determined how each event should be taped given the procedures we had developed.

Once we had a description of how the test lesson should have been taped, we evaluated each videographer's performance against this ideal. We used a three-point scale to score how well they taped each of the 22 lesson events. The videographers were given a score of zero if they broke any of the rules that they needed to take into account. For example, if they did not zoom in to capture information that they were supposed to capture, or if they pointed the camera at the wrong thing, they would be given a score of zero. They were given a score of one if they showed an understanding of the rule they needed to carry out but did not apply it in a timely fashion. For example, if they needed to zoom in and capture what the teacher was pointing to but reacted too slowly and missed this information, or if they let the teacher walk around the class for a while before they decided to follow her or him, they would receive a score of one. They were given a score of two if they applied the rules exactly as we had predicted they should.

The scores obtained were all in a similar range and also were relatively high. The German videographer received a score of 35 out of a possible total of 44. The Japanese videographer received a score of 36, and the American videographer a score of 43. In addition, of the 66 events scored for the three videographers, only 4 were rated a zero (which means that a rule was actually broken only 4 times). Two of these zeroes were obtained by the German videographer, and two by the Japanese videographer. This means that no videographer ever showed more than two rule breaches for the entire test.

Performance on the test was also used to evaluate the quality of each videographer's camera work. First we generated a list of possible flaws that a videographer might produce. Our list included the following flaws:

- Cropping shots too tightly (e.g., cutting off part of someone's head).
- Cropping shots too wide (e.g., too much head room).
- Zooming in/out and then having to reframe the shot.
- Zooming in/out and then having to refocus the shot.
- Panning while zoomed in tightly.
- Jerky or awkward camera movement during zooms or pans.
- Losing from the frame any object that is being tracked.
- Unnecessary camera movement.
- Bad coordination between zooms and pans.
- Very unbalanced composition.

We used this list to score each videographer's performance on a four-point scale for each of the 22 events in the test lesson. Videographers were given a score of three on an event if we could

find no flaw in their camera work. They received a score of two if one flaw could be found, a score of one if two flaws could be found, and a score of zero if at least three flaws could be found.

All videographers obtained scores that were within a similar range and judged to be satisfactorily high. The Japanese videographer received a score of 51 out of a possible total of 66. The German videographer received a score of 52, and the American videographer a score of 60. Both evaluations of the test confirmed our informal impression that camera standardization had been reached by the end of the training.

Videographers were in the field for a prolonged period of time. We worried, therefore, that they might slowly forget what they were taught or develop bad habits. In order to make sure that they continued using the camera correctly, every 10th tape that came in from the field was evaluated using a scoring system similar to the one described above. Videographers were given feedback about how they were doing. In particular, they were immediately informed if they had, in any way, drifted away from the standards we knew they were able to follow.

Gaining Cooperation from Teachers

We were concerned at the outset of the study that we would have difficulty finding teachers who were willing to be videotaped. Anticipating such difficulty, we decided to pay teachers for their participation. However, our fears may have been unfounded. In fact, getting schools to participate in the main TIMSS study proved to be more difficult than getting them to participate in the video study. I believe this is because the actual demands imposed by videotaping are minimal compared to those imposed by testing of students. As video surveys become more commonplace, it may prove easier and easier to secure cooperation from teachers, so long as videotaping is not tied to accountability for individual teachers.

Some Notes on Equipment

The quality of the data depends to a great extent on the quality of the equipment used in collecting the data. Thus, we wanted to use high-quality cameras that would produce excellent images, and high-quality microphones that would enable us to hear most of what goes on in the classroom.

The camera we selected was a Sony EVW-300 three-chip professional Hi-8 camcorder. Each camera was mounted on a Bogen fluid-head tripod. (Tripods that are not fluid head will produce jerky camera movements.) A small LCD monitor was mounted on the camera to help operators view what they were taping. Sound was collected using two microphones, one a radio microphone worn by the teacher, the second a shotgun zoom microphone mounted on the camera. Good audio is both difficult to achieve in classrooms, and extremely important for analyzing the contents of the tapes. Thus, it is best to purchase the highest quality microphones available.

Constructing a Multimedia Database

As the tapes and supplementary materials are collected, they are mailed to our project headquarters at UCLA. The tapes are then processed as follows: Videotapes and supplemental images are digitized, compressed, and stored on CD-ROM. Using software we have developed for this study, videotapes are transcribed, translated into English, and marked with time codes so that transcripts and video can be linked in a multimedia database. In the following sections we will describe these procedures in more detail.

Digitizing, Compression, and Storage on CD-ROM

The first step in constructing the multimedia database is to store the videotapes and supplementary materials in digital form on CD-ROM disks.

Because video contains so much information, it has until recently not been feasible to store large quantities of video in digital form. The breakthrough that makes such storage possible has been in the development of algorithms for compressing digital video so that it can be stored in smaller and smaller spaces. The algorithm we are using in the current project is called MPEG-1, an algorithm endorsed by the Motion Picture Engineers Group, that is fast becoming the industry standard. MPEG compression makes it possible to store 74 minutes of video and audio on a single CD-ROM disk.

Once we receive our videotapes, we digitize the tapes and compress them into an MPEG file on a large hard disk. Text pages, worksheets, and other supplementary materials collected by the videographers are digitized on a flatbed scanner and stored in PICT format on the same hard disk drive as the accompanying videotape.

Once the MPEG file and accompanying PICT files for each lesson are stored on the hard disk drive, the files are burned onto a CD-ROM.

Software and Hardware for Accessing Digital Video

Once the video is stored on CD-ROM disks, it can be accessed by the database software we have developed for this project. Users of the software work at a computer workstation consisting of the following:

- Apple Macintosh Power PC 8100AV computer with built-in CD-ROM drive;
- Apple 17-inch Multi-Scan monitor;
- Hardware card in computer for real-time decoding MPEG files (manufactured by Wired, Inc.); and
- Headphones.

Workstations are networked together in a client/server system. The server consists of a Macintosh Power PC 8100 computer. Although video is stored locally on CD-ROM at each workstation, all transcription/translation and time codes that link the transcription to the video are stored on a central server. This makes it possible for many transcribers and coders to work simultaneously on a single, integrated database. It also means that later, in the analysis phase, we will be able to apply sophisticated search procedures to the entire database at once, without having to change CDs. Only if we need to view the video itself will it be necessary to locate and load the actual CD.

We have so far implemented three modules in the software: transcribe, code, and analyze.

The transcribe module enables transcribers/translators to:

- View the video and control playback through a window on their computer screen;
- Type the transcription/translation into another window on the screen; and
- View the video, once transcribed, with subtitles in real time.

The transcriber sees two major windows on the computer screen: one displays video, the other displays the transcript. Under the video window is a rectangular area used for displaying transcript records as subtitles in real time, and various buttons for controlling the video. Various controls allow the transcriber to:

- Set up and easily modify a continuous loop so they can watch the same segment of video over and over while they transcribe/translate the speech;
- Move the loop forward to continue transcribing the next segment of video;
- Stamp time codes to mark the beginning of each utterance;
- Enter new records into the transcription database;
- Merge records together and break records apart;
- Move instantly to the point in the video that corresponds to the highlighted transcript record;
- Move instantly to the point in the transcript that is closest in time to the point where the video currently is; and
- Turn synchronized subtitles on and off while viewing the video.

Transcription/Translation of Lessons

Our goal is to have transcripts that reflect, as accurately as possible, the words spoken by both the teachers and the students. It is not enough to summarize or paraphrase the talk, nor is it acceptable to transcribe the data in a way that reflects what the participants mean to say.

We have developed a protocol to make sure that all transcription/translations are carried out in a standardized manner. For example, transcribers are given rules about how to indicate speakers,

how to break speech into turns, how to use punctuation in a standardized manner, and how to translate technical terms in a consistent way.

Each American lesson is transcribed in order to facilitate coding. Because some parts of the video are hard to hear, the transcript enables the coder to better understand what is happening in the lesson. It also is possible to code some aspects of instruction directly from the transcript, without viewing the video at all.

German and Japanese lessons are translated into English as they are being transcribed. The purpose of the translation is to aid in multilanguage searches of the database, and to make it possible for persons not fluent in German or Japanese to view and understand the lessons. All coding of the videotapes will be done by native speakers of the language being coded. Thus, coders will not rely on translations to make subtle judgments about the contents of the video.

Videotapes are transcribed and translated by teams of transcribers fluent in each of the three languages. Some members of the German and Japanese teams are native speakers of those languages, others are native speakers of English but fluent in German or Japanese. Each tape is transcribed/translated in two passes. One person will work on the first pass transcription/translation of a tape, and then a different person is assigned to review this work. A hard copy of the first pass transcription/translation is printed out, and the reviewer marks any points of disagreement on this copy. The two individuals then meet, discuss all the proposed revisions, and come to an agreement about what the final version should be. In cases where disagreements cannot be resolved, a third party is consulted.

The last step in the transcription/translation process is to time code the tapes, i.e., to mark the exact point at which each utterance begins.

Coding and Analysis

Instructional quality is a complex construct for which few standard indicators exist. Coding of classroom videotapes, therefore, is part of a cyclical process that involves refining the construct, developing indicators of the construct, validating the indicators, and then using the results to further refine the construct. The state of the art of this process is at a very rudimentary level: we have poor ways of describing classroom processes at present. Partly this is because classroom instruction is a highly complex system that is inherently difficult to describe. It is also true that we have devoted far less energy to this enterprise than to measuring the outcomes of instruction.

This section will provide a description of how we began to develop the coding system for the TIMSS video study, and how we are implementing the coding in our LAVA software program.

Deciding What to Code

In deciding what to code, we had to keep two goals in mind: first, we wanted to code aspects of instruction that relate to our developing construct of instructional quality; second, we wanted the codes we used to provide us with a valid picture of instruction in three different cultures. For the first goal, we sought ideas of what to code from the research literature on the teaching and learning

of mathematics, and from reform documents—such as the *NCTM Professional Teaching Standards*—that make clear recommendations about how mathematics ought to be taught. We wanted to code both the structural aspects of instruction, i.e., those things that the teacher most likely planned ahead of time, and the on-line aspects of instruction, i.e., the processes that unfold as the lesson progresses.

The dimensions of instruction we judged most important included the following:

- *The nature of the work environment.* How many students in the class? Do they work in groups or individually? How are the desks arranged? Do they have access to books and other materials? Is the class interrupted frequently? Do the lessons stay on course, or do they meander into irrelevant talk?

- *The nature of the work that students are engaged in.* How much time is devoted to skills, problem solving, and deepening of conceptual understanding? How advanced is the curriculum? How coherent is the content across the lesson? What is the level of mathematics in which students are engaged?

- *The methods teachers use for engaging students in work.* How do teachers structure lessons? How do teachers set up for seatwork, and how do they evaluate the products of seatwork? What is the teacher's role during seatwork? What kinds of discourse do teachers engage in during classwork? What kinds of performance expectations do teachers convey to students about the nature of mathematics?

Our second goal was to accurately portray instruction in Germany, Japan, and the United States. Toward this end, we were concerned that our description of classrooms in other countries make sense from within those cultures, and not just from the American point of view. One of the major opportunities of this study, after all, is that we may discover approaches to mathematics teaching in other cultures that we would not discover in our culture alone. We wanted to be sure that if different cultural scripts underlie instruction in each country, we would have a way to discover these scripts.

For this reason, we also sought coding ideas from the tapes themselves. In a field test, we collected nine tapes from each country. Collected in May 1994, we convened a team of six code developers—two from Germany, two from Japan, and two from the United States—to spend the summer watching and discussing the contents of the tapes in order to develop a deep understanding of how teachers construct and implement lessons in each country.

The process was a straightforward one: we would watch a tape, discuss it, and then watch another. As we worked our way through the tapes, we began to generate hypotheses about what the key cross-cultural differences might be. These hypotheses formed the basis of codes, i.e., objective procedures that could be used to quantitatively describe the videotapes. We also developed some hypotheses about general scripts that describe the overall process of a lesson, and devised ways to validate these scripts against the video data.

Developing Coding Procedures

Once the list of what to code has been created, we are ready to begin developing the specific procedures to be used in coding the tapes. First, field-test tapes are viewed by the coding development group, and a definition of the category to be coded is proposed. Then, code developers try to apply the definition to the field-test tapes from their country. Difficulties are brought back to the group, and definitions are revised and refined. This process is repeated until all members of the group are satisfied with the definitions and procedures, and agree with the coding of each instance.

Once codes are developed, coders are trained to implement the codes. Before coding begins, a formal reliability assessment is conducted to ensure independent agreement across coders at a level of at least 80 percent for each judgment. Reliability is assessed by comparing each coder's results with a standard produced by the coding development team.

Throughout this process we endeavor to be strategic. For example, just having collected 100 hours of video does not mean that all 100 hours must be analyzed. Depending on the frequency of what is being coded, it may be possible to time sample or event sample, and our computer software makes this easy to do. It is also important to divide coding tasks into passes through the data in order to lessen the load on coders. This increases reliability and speeds up coding.

Implementation of Codes Using the Software

The code module of our software enables coders to view synchronized video and transcript on their computer screen. On-screen controls allow them to move instantly to the point in the video that corresponds to the highlighted transcript record, or to the point in the transcript that is closest in time to the current frame of video.

Coders can work from video, transcript, or both, and they can mark the occurrence of events they are targeting in a given coding pass.

There are three types of events that can be coded:

1) In only—an event is marked by a single time point. Events would be coded this way when we do not care to measure their duration but just want to record their occurrence.

2) In and out—an event is marked with a beginning and end point on the videotape. Most of the events we code are of this type. For example, we code when periods of seatwork begin and end.

3) Exhaustive segmentation—a tape is segmented such that the end point of one segment serves as the beginning of the next, meaning that no part of the tape is not included in a segment. We use this type of event when coding classroom organization, for example. Coders are forced to categorize each part of the tape into one of the three categories of organization.

The software enables coders to code events from video by marking a beginning and ending, or beginning only, time code; or from transcript by marking the beginning and ending, or beginning only, points in the transcript. It also allows us to define new event types by searching Boolean combinations of other events and characteristics that have already been coded.

The software also allows the coder to characterize an event that has been coded. A button on the screen takes coders to the next event that has been coded, plays the event, and then presents the options for coding of characteristics. There are four types of characteristics that can be coded.

1) Numerical—an event is characterized by a numerical value on some dimension.
2) Mutually exclusive—an event is categorized into one of a mutually exclusive and exhaustive set of categories.
3) Check all that apply—an event is judged as belonging to one or more of a set of non-mutually exclusive categories.
4) Descriptive—a qualitative description is written and attached to a particular event.

Codes can be applied using one of four sampling schemes.

1) Play all—the coder can watch the entire lesson, marking codes whenever they are appropriate.
2) Play events—the coder can watch only events of a particular type, then characterize the events.
3) Sample events—the coder can be presented with a randomly chosen sample of events of a particular type.
4) Sample time—the coder can be presented with a randomly chosen sample of time segments, then mark whether or not specific events happened during each segment.

First-Pass Coding: The Lesson Tables

We have found that it is useful to have an intermediate representation of each lesson that can serve to guide coders as they try to comprehend a lesson, and that can be coded itself. For this purpose, our first step in coding the lessons is to construct a table that maps out the lesson along the following dimensions:

- Organization of class—each videotape will be divided into three segments: pre-lesson activities, lesson, and post-lesson activities. The lesson needs to be defined in this way because the lesson will be the basic unit of analysis in the study.

- Organization of interaction—the lesson is divided into periods of classwork and periods of seatwork.

- Activity segments within classwork—each classwork segment will be further divided, exhaustively, into activity segments according to changes in pedagogical function. We have identified seven different kinds of activity segments: introduction, instructing,

setting up seatwork, sharing seatwork product, correcting homework, test-taking, and conclusion.

- Activity segments within seatwork—we have distinguished three types of activities during seatwork: working on tasks and situations, correcting homework, and correcting seatwork. In addition, we have added two categories to characterize the kinds of simultaneous activities we have seen thus far: working and correcting homework, and working and correcting seatwork.

- Mathematical content of the lesson—the mathematical content of the lesson is divided into units. The content of each unit will be written down concretely/qualitatively, and then categorized into one of four types: situation, task, information, and solution method.

We are using these first-pass tables for two purposes. First, they can be used by subsequent coders to get oriented to the contents of the videotapes. Often it takes a great deal of time for coders to figure out what is happening in a lesson. The tables ease the way, providing an overview of the structure and content of each lesson.

A second purpose for the tables is that some codes can be coded from the tables without even going back to the videotapes. Examples of such codes include TIMSS content category, nature of tasks and situations, and changes in mathematical complexity over the course of the lesson.

Confidentiality and Sharing of Data

As pointed out above, there is a major issue concerning how to make video data available for secondary analysis while at the same time protecting the confidentiality of study participants. We have outlined one approach to accomplishing these goals as part of a proposal to establish the TIMSS video data as a restricted-use data set.

Our strategy for preserving the confidentiality of participants will be similar for both raw and restricted-use data sets. In general, we will separate the activity of coding the visual images (e.g., access to video pictures of teachers and students) from the activity of analyzing the results of the coding. Persons engaged in coding will have no access to any identifying information about teachers or students. They will know which country the teachers are from, but nothing else. Persons engaged in analysis, on the other hand, will work with data sets in which summary variables from the coding have been linked, via a teacher ID, to other information from TIMSS. But these analysts will not have access to video images.

This will be accomplished by constructing two independent data sets, one for the video data, the other for all other data. Separate ID numbers will be assigned to teachers in each data set. Information that can match IDs from one data set to the other will be held in a secure place, available only to senior personnel. A third, integrated data set will be constructed once we are ready to undertake integrated analyses. This integrated data set will not contain any visual images.

For the restricted-use data set, additional safeguards would be taken to make it practically impossible for researchers to link the two data sets with identifying information.

First, all specific identifying information would be deleted from the second data set; researchers would be provided with only a subset of variables that were available in the raw data set. For example, geographic region of the country would be deleted, as would size of school, age of teacher, and so on.

Second, we would exercise controls over the coding of video data that would prevent researchers from linking any specific image with any other data, although codes, of course, would be linked. We propose using the following procedures:

- Access to video data would be allowed only in specifically designated research rooms in which the full data set would be available. Researchers could view and code video data in this room or rooms, but would have no access to the second data set at all while they were coding video data. Researchers would not be allowed to remove any written materials from this room.
- After researchers complete their coding of the video images, project staff would construct aggregate data sets containing the results of the coding, remove all ID numbers, and then give the data back to researchers in an electronic spreadsheet format for analysis. Researchers who wanted additional TIMSS data integrated into their video coding spreadsheet would simply request that project staff put the additional variables into their spreadsheet. Again, all ID numbers would be deleted.

We believe that these safeguards would provide a high degree of confidentiality to participants while at the same time allowing researchers to access this valuable and unprecedented data set. Of course, if a researcher brought up an image and said “Oh, that's my sister-in-law,” confidentiality would be undermined. But such an event is unlikely.

CONCLUSION

I began this paper by urging a new emphasis on developing and using observational indicators of classroom processes. I proposed video surveys as a promising approach to this task, but outlined some difficult issues in the implementation of video surveys. Finally, I showed how, in the TIMSS Videotape Classroom Study, we have successfully resolved these issues in the first large-scale video survey of classroom instruction.

Although I believe I have shown that video surveys are logistically possible, it is too early to see what the full benefits of such studies will be. The technology for assessing student outcomes has been developed over a long period of time. Research on classroom processes, in contrast, is still in its infancy. There is much work to be done before statistically acceptable, useful indicators are in hand. The task of developing such indicators, however, strikes me as one of the most important to be undertaken over the next decade. If we cannot make significant progress on the assessment of instructional processes, we will not have the basis on which to improve classroom instruction. Without this solid empirical foundation, efforts to reform instruction will continue to be grounded in ideological debates and pendulum swings.

NOTES

1. A more detailed account of these procedures can be found in the “TIMSS Videographers' Handbook,” available by request from the author.

REFERENCES

Cronin, J.M. 1985. "Issues in National Educational Data Collection," NCES Redesign Project.

Hall, G., Jaeger, R. M., Kearney, C.P., and Wiley, D. E. 1985. *Alternatives for a NationalData System on Elementary and Secondary Education.* Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement.

Husen, T. 1967. *International Study of Achievement in Mathematics.* New York: Wiley.

McKnight, C.C., Crosswhite, F.J., Dossey, J.A., Kifer, E., Swafford, J.D., Travers, K.J., and Cooney, T.J. 1987. The Underachieving Curriculum: Assessing U.S. School Mathematics from an International Perspective. Champaign, IL: Stipes.

Peterson, P.L. 1985. "The Elementary/Secondary Redesign Project: Assessing the Condition of Education in the Next Decade," NCES Redesign Project.

Porter, A.C. January-February 1995. "The Uses and Misuses of Opportunity-to-Learn Standards," *Educational Researcher* 24 (1): 21–27.

Discussant Comments

KEVIN F. MILLER

NCES has supported developing a new technology that offers the promise to revolutionize our understanding of the processes that go on in classrooms and, in turn, to dramatically increase the impact of the research NCES supports. In this comment, I will 1) describe some of the consequences of this new technology; 2) discuss some aspects of human cognition that make it particularly important; and 3) argue that NCES could play a pivotal role in creating a new American Education Yearbook, including a video archive of educational processes in American schools.

Videosemantics: Making Sense Out of Classroom Processes

NCES collects data on teachers and classrooms as a method of describing the changing face of instruction in the United States and making it possible for researchers and policymakers to understand the instructional processes that account for changes in educational achievement. This is primarily done through surveys of teacher's beliefs, attitudes, and activities. As Stigler (1995) notes, there are fundamental problems in moving from these data to a real understanding of what goes on in classrooms. Self-reports of teaching practices may not produce accurate descriptions of actual classroom processes, because teachers may vary in how they interpret survey questions and may have limited and selective recollection of what transpires in their classrooms.

There is a more fundamental obstacle to going from surveys to prescriptions for improving instruction. In the same way that knowing the ingredients in a cake does not by itself enable you to bake one, knowing the characteristics of a good teacher does not in itself tell you how to become one.

What is needed to move from descriptions to prescriptions is a method of making the process of instruction explicit, and this is precisely what the video survey technology provides. The actual process of instruction can be made accessible to scientific study in a way that has been hitherto impractical. Observers could easily watch how 20 different teachers teach the same content, or how the same teacher responds to the questions of different students. Teachers in training could observe how skilled teachers respond to problems that come up in the course of instruction, and could watch themselves as they attempt to teach a lesson.

In his paper describing this technology, Stigler (1995) notes that the key to the revolutionary improvement in manufacturing quality engineered by W. Edwards Deming was the insight that

improving quality requires one to focus on the processes of manufacturing rather than simply inspecting the products of those processes. NCES has made possible the development of technology that could lead to a similar revolution in education, by changing the focus from testing students and surveying teachers to actual measurement and description of instructional processes.

Taming the Power of the Anecdote

Video technology may also provide a solution to one of the most vexing problems facing educational researchers: the enormous difficulties that the consumers of research have in understanding statistical data. People are much more likely to be swayed by individual anecdotes than they are by carefully collected, representative data. A good demonstration of this problem was provided by Borgida and Nisbett (1977), who presented University of Michigan psychology undergraduates with evaluative information about upper level courses in their field. This was either presented by previously unknown confederates as representing their personal experience, or as the ratings of an entire group of students. Despite what these students must have learned about the effect of sample size on the reliability of observations, the individual reports had a significantly larger impact than did the statistical data on whether or not students planned to sign up for the recommended courses and avoid the non-recommended courses.

At its most extreme, the power of the anecdote that suggests the pictures appearing on the cover of an NAEP report may have more impact than the data contained inside. Statisticians may bemoan the power of anecdotal experience, yet it appears to be a fundamental aspect of human cognition. Video technology offers a potential solution here, providing a means for turning vignettes into data that can be presented systematically. Observers can code a corpus of classroom observations, producing quantitative descriptions of the data set. These quantitative descriptions can be coupled with presentations of examples of the kinds of processes observed. Because these observations are culled from a data set, it is possible to determine whether they are representative or exceptional, and it is possible for researchers with different interests to code the same data set in different ways. The melding of statistics and anecdotes that the video technology makes possible can be both powerful and methodologically responsible—powerful in the way that only direct experience can be, and responsible in that the statistical representativeness of these experiences can now be assessed.

Exploiting the Technology: A Yearbook of American Education

NCES has supported the developing of a revolutionary method for collecting educational data and making it accessible to researchers. It has an equally vital role to play in promulgating this technology and ensuring that it is used to understand the changing state of instructional processes in the United States. Imagine how valuable it would be if there were a systematic filmed record of teaching in the United States from earlier eras. Such a database would be a gold mine for researchers interested in all aspects of changes in the lives of children and the processes of education. A database of current instruction in the United States will be equally valuable for anyone who wishes to understand the changing face of schooling in America. NCES has experience in and sampling of the state of education in the United States. It should be within both its expertise and mission to develop a video yearbook of American education by collecting a representative sample of teaching in the United States. Such a database would be of interest to researchers and policymakers from a

variety of fields. It would not only provide a vital record of the state of American education but also would be extremely useful in helping us to understand the classroom processes that result in effective instruction. Additionally, it would form a lasting legacy for future generations, who will use it to answer questions that we cannot now anticipate.

References

Borgida, E., and Nisbett, R. E. 1977. "The Differential Impact of Abstract vs. Concrete Information on Decisions." *Journal of Applied Social Psychology* 7: 258–271.

Stigler, J. W. 1995, November. "Large-Scale Video Surveys for the Study of Classroom Processes. Paper presented at the NCES conference "Future NCES Data Collection: Some Possible Directions." Washington, D.C.

8 Education for Work: Curriculum, Performance, and Labor Market Outcomes

THIS PAGE INTENTIONALLY LEFT BLANK

Education and Work: Curriculum, Performance, and Job-Related Outcomes

Peter Cappelli

EXECUTIVE SUMMARY

Perhaps the most fundamental question within the topic of education and work is whether the two are in conflict. Are the requirements for success in the workplace in conflict with the goals of academic achievement? Putting this issue to rest would be an enormous contribution, but it requires data of the kind outlined below.

CHANGES IN THE WORKPLACE

Evidence suggests that skill requirements are clearly rising for many jobs, perhaps for the average, but not uniformly. The skills that are in increasing demand are often the kind of behavioral skills that have not typically been part of academic achievement assessments.

Declining attachment between employer and employee raises questions as to where workers will get skills. It puts the burden more on the education system, as we should expect people to go back and forth from school and work, repeating some of the school-to-work transition issues over a lifetime.

WHAT DOES WORK DEMAND OF EDUCATION?

We do not really know the answer to this question because most exercises simply ask employers whose requirements are always in flux. We need ways to validate estimates of the effects of education on actual job and organizational performance.

To do this, we need better data in three areas:

1) Identifying the knowledge, skills, and abilities (KSAs) produced by education, especially those traditionally excluded in academic achievement assessments;

2) Identifying the characteristics of schools and education that produce the desirable KSAs; and

3) Measures of performance in the workplace that go beyond wages in order to examine both the success of individuals and their organizations.

HOW DOES WORK AFFECT EDUCATION?

Longitudinal data are needed that go beyond the simple cross-sectional studies aimed at secondary school academic achievement, especially as we are increasingly concerned with lifelong learning. Again, we need better data in three areas:

1) Data and analyses relating work experience to postsecondary achievement, broadly defined;
2) The effects of work on a wide range of learning, including work-based skills, behavioral skills, and so on; and
3) More complete information on work experience, including the nature of tasks performed and the learning experiences at work.

SPECIFIC IMPLICATIONS FOR DATA COLLECTION

We need to link information about work experiences and education experiences in the same data sets, as well as more thorough measures of inputs and outcomes for both education and work in these data sets. The best approach is to leverage off of existing data sets as follows:

- Additional information on work experiences could be added to the NLS–72 and HS&B data sets;
- New longitudinal surveys are needed to collect more data on educational experiences and outcomes, especially for secondary school; and
- Data on employers represents the biggest challenge in order to understand how employer practices affect later education and how education and KSAs, in turn, affect organizational performance. More targeted surveys that match employers and employees might be the most cost-effective approach.

It has long been understood that education has an important influence on success in the workplace. More recently, many observers believe that this influence is becoming more important and that the benefits of education may well extend beyond the success of an individual worker to that of organizations and entire economies. With this visible change has come increased interest in exactly how education affects workplace performance. For example, does the subject matter and the pedagogy used affect workplace outcomes in addition to the credentials one attains, and are there innovations that could be made in the education system that would strengthen the relationship between education and performance?

The potential effects that work can have on education, on the other hand, have perhaps been less appreciated outside of the research community. The interest in work-based learning and the

identification of skills that are best learned in the context of actual workplace experience are examples of the type of effects that work can have on education and learning.

The education research community has not always been especially interested in relationships with work. The understanding of the role education plays in labor market success, for example, was identified and championed by labor economists who were interested in understanding wages; the role that KSAs conveyed in education play in determining job performance was identified and researched by personnel psychologists whose goal was understanding effective employee selection; and the importance of work-based skills and learning has largely been advanced by studies of international competitiveness that emphasized the role that apprenticeships and other school-to-work programs play in raising national skill levels. The effects of work on traditional education have been perhaps more thoroughly examined by education researchers, although here the focus has often been either negative (linking student hours of employment to poor academic achievement) or highly focused on legislated programs such as review of vocational education programs mandated by the Perkins Act.

Perhaps the main reason for the relative lack of interest—and in some cases antipathy—in the workplace among the community of education practitioners and scholars was the sense that what mattered for workplace success was different, perhaps even antithetical, to the factors that shaped academic success. Scholars like Bowles and Gintis (1976) asserted that what employers wanted and perhaps needed from schools in terms of the characteristics of graduates/new hires was a kind of compliant behavior that was in conflict with the goals that educators held for their students. There are certainly arguments and evidence suggesting that they are not in conflict, but the view that they are remains deeply held in many circles.[1]

One very important consequence of this perception of conflict in goals has been continued support for an inward orientation toward evaluation in education: The "success" of an education establishment, for example, has almost uniformly been based on how well its students learn the material that educators have presented as assessed by the education community itself or, at the secondary school level, on how well graduates do in getting access to postsecondary education. Whether learning that material contributes in some important way to other life outcomes is rarely examined. Consider, for example, what the equivalent arrangement sounds like in a different situation like medicine. Procedures would be evaluated based on whether they did what the doctors wanted them to do and not necessarily whether it furthered the patient's health. "The treatment was a success but the patient died" is the aphorism used to parody such arrangements in medicine.

One of the first general priorities for NCES and the research community should be to address whether the goals of educators for students are in fact in conflict with the goals of workplace success. Specifically, whether achieving in school based on traditional measures is related to or in any way in conflict with achieving in the workplace. If the perception of conflicting interests can be put to rest, then at least some of the conflict between business and education may abate as well. This will also apply to at least some of the resistance to evaluation based in part on workplace outcomes in the education community, as well as the lack of real participation and commitment to education among the employer community. If, on the other hand, conflicts are identified between these goals,

such information would provide important evidence for striking compromises or creating new arrangements for advancing success in both arenas.

Several factors are pushing the education community away from internal assessments and toward evaluations that are based more on external criteria. These efforts are widespread, and some are likely to be much more productive than others. On the negative side, they include pressures for "accountability" in the public sector, which have played out in postsecondary education as efforts to judge the efficiency of state education systems in crude cost-benefit terms, graduates per dollar. Attention to the workplace success of school leavers as part of the assessment of education is a potentially more useful development. It has been powerful for several reasons.

First, employers and policy observers have been vocal in their belief that the poor preparation of school leavers has contributed to problems inside organizations and in the economy as a whole (the extent to which they are right in thinking so is another matter). The reports making these arguments are so well known as to be almost household names. They include *A Nation At Risk* (1983), *Workforce 2000* (1985), and *America's Choice* (1989). The legislation that resulted in part from these arguments could institutionalize the interest in relating work to education, subject, of course, to continued funding from Congress. These include the School-to-Work Opportunities Act of 1994, with its efforts to develop infrastructure at the state level to bring school and work closer together, and the mandate of National Goals for Education Act of 1992 to develop national skill standards for jobs that can then be translated into curricula and credentials for participants.

Second, as the job market tightens, students and their parents will increasingly demand that schools—primarily postsecondary institutions—do a better job in preparing students for the workplace. Entry-level wages for college graduates have been falling rapidly in real terms— much more so than for the work force as a whole—while the proportion of college graduates who will find jobs requiring college skills is projected to decline.[2] The anecdotal reports from state university systems that as many as 25 percent of 4-year graduates return to community colleges for work-based classes before getting a job suggests something about the magnitude of the problems in preparing students for the workplace.

CHANGES IN THE WORKPLACE[3]

Behind the above pressures on education are profound changes in the workplace that will make very different demands on education systems and, more to the point, increase the importance of education suppliers to workers and the economy. The first of these is the change inside organizations as to how work is organized. Specifically, what new tasks are workers required to perform, and what different skills do those tasks demand from them?

Whether new models of work organization are in fact changing skills—and, if so, in what way—is a central question for advocates who believe that we need different kinds of data for research. Whether skill requirements are a more important issue now, where the kinds of skills that

are important have changed, and whether these skills challenge existing data collection efforts are among the issues driven by workplace changes.

The place to begin that discussion is by asking what is happening, on average, to job requirements. Are skill demands really changing as much or in the manner that many advocates suggest, creating real shortages of workers with the education level necessary to fill jobs? The recent EQW/Census National Employer Survey (1995) found a majority of employers asserting that overall skill requirements have risen in their organizations for production or front-line jobs. But it may not be obvious how valid these responses are given the subjective nature of the responses where "skill" is not defined, for example, and can easily be confused with performance requirements. In other words, more may be demanded from employees, but what is really being increased is effort, not skill.

We used the EQW survey to examine what factors seem to differentiate those establishments reporting that skill requirements have risen for their front-line workers (Cappelli 1995). Those that have Total Quality Management Programs (TQM), more extensive teamwork arrangements and greater use of computers for both managers and non-managers, report that skill needs are increasing. These changes are consistent with the arguments that the shift toward "high-performance" workplaces is raising the skills needed in establishments that introduce those practices. As these practices become more widespread, these developments could have economy-level consequences. Establishments with more educated workers are also more likely to report that skill requirements are rising. This result is consistent with the arguments made by Bartel and Lichtenberg (1987) that more educated work forces have a comparative advantage in adopting innovations in technology and practices that might raise skill needs.

A different approach might be to look within establishments at the actual changes going on in the way work is organized. Consider, for example, the issue of autonomy, a key concept in participative work systems and an important factor in raising skill requirements. The argument is that as participative and decentralized work systems expand, employees have much greater autonomy in decision making and therefore need much greater skills to make the kind of decisions that their more highly trained supervisors had made for them in the past. But as Klein (1989) observes, just-in-time inventory systems that eliminate buffers of materials or intermediate products between work groups make those groups highly interdependent; changes in the production arrangements within any individual group can change its work pace, causing either shortages or pile-ups of material downstream. Because the overall flow of work across *all* teams in the assembly process must be absolutely consistent, the autonomy that any individual worker or team has to make changes in work organization is tightly constrained.

Further, as Adler (1993) discovered at the New United Motors (NUMMI) joint venture between Toyota and General Motors, the principle of continuous improvement requires that the performance of individual tasks be completely routinized so that the work teams can discover whether minute changes in tasks lead to an improvement in performance. In this sense, continuous improvement in work processes is like a laboratory experiment where everything is held constant except the one change being investigated. For employees, individual tasks appear to be every bit as rigidly defined as under scientific management. Individual workers in fact do not have the kind of autonomy that demands higher skill levels. The fact that the work teams themselves can influence

the design of those tasks may make the system more palatable, however. In manufacturing, therefore, where most of the reform efforts have been concentrated, innovative production processes may not necessarily lead to work organization that makes dramatically different demands on production employees.

My study of changes in skill requirements used data obtained on 56,000 production workers over an 8-year period to examine whether skill requirements have changed. The results suggest significant upskilling for production jobs across the board as measured by changes in Hay points, the job evaluation metric used by Hay Associates to measure job requirements. Some of the upskilling seems due to the fact that tasks associated with quality control and housekeeping have been pushed onto all the remaining jobs (the decline of employment in quality and housekeeping jobs is consistent with this interpretation). That is, not only has each job experienced upskilling but also the overall distribution of production jobs has shifted away from less skilled and toward more skilled positions (Cappelli 1993).

"Lean production" techniques that have become popular in manufacturing (see below) essentially eliminate some jobs and push their tasks onto production workers. Some of those tasks, such as housekeeping, add little to the job. Other tasks, such as coordinating job design changes across teams, demand considerably higher skills, especially behavioral skills (communication, negotiation, and group dynamics skills). Adler (1993) notes that many of the tasks previously performed by industrial engineers, such as job analysis and redesign, are now being pushed down to the production teams.

It is also important to remember that while these skill requirements are rising, they start at a low base. Data from Hay Associates suggest that a typical management job, for example, has skill levels about twice those represented by production work. Given the low base, it is certainly possible that workers already have the skills to meet the increasing skill demands represented by these data. In other words, the fact that job requirements are rising does not necessarily mean that workers' existing skills are likely to be challenged.

Is There "Upskilling" Outside of Production?

By definition, the techniques of high-performance *production* systems are associated with production work, and not all of these techniques apply directly to other industries. The equivalent study to the one noted above using Hay data for clerical jobs finds no consistent pattern; some clerical occupations show increases in skill while others experienced decreases (Cappelli 1993).

One important attribute of the "lean-production" or "high-performance" work systems that do seem to raise skill requirements in manufacturing is the increased flexibility needed to handle variations in products. Situations that do not demand change—indeed may punish it—may not make great use of these techniques. There is relatively little use of high-performance production techniques in industries like transportation, distribution, or public utilities, perhaps because reliability and consistency are the prime considerations there. Indeed, the work systems in these industries are often referred to as "high-reliability" systems.

One of the more curious findings, however, is that there is little evidence of work practices associated with high-performance production systems even in organizations that have production-

like aspects. The processing of transactions in the back offices of financial services and related industries, for example, looks very much like an assembly line (more people are employed in these industries than in manufacturing). Yet there appears to be little—if any—evidence that high-performance production practices or even specific high-performance work practices are being used in these operations. Indeed, the effort in these facilities seems to be quite strongly in the opposite direction; to automate employees out of the process altogether.[4]

It is not obvious that there is a common trend in service jobs. In health care, for example, anecdotal evidence suggests that the biggest development has been efforts to deskill jobs along the lines of Taylorism: Many of the simple tasks traditionally performed by nurses are now being transferred to lower skilled workers. In customer contact jobs in retailing and hospitality, there are some efforts to "empower" workers by giving them more authority to solve problems. Overall, there appears to be a clear trend toward high-performance work in production-oriented jobs because it is associated with a new production process. It is not clear that this movement will make the same inroads elsewhere.

What Skills Have Changed?

In situations where new work practices are in place, how have the jobs changed? Consider, for example, the tasks transferred to work teams in high-performance work systems in a manufacturing environment. The systems of performance measurement and control are already in place, as is the existing job design. The task facing the teams is simply to learn how to interpret information from the system in order to look for ways to improve it. They are not designing and setting up a new system. Further, because these decisions are made in teams, it is not necessary for each worker to have all of the skills needed to handle every task, only that those skills be available somewhere in the work group, perhaps spread across different individuals. For example, not every worker in the group needs to understand how to use statistical process control techniques. If one person understands the notion of confidence limits, another can read the charts, and a third knows his or her machine tools well enough to troubleshoot when the problems have been identified, they have a team that can make the technique work.

Another study examines the relationship between these new work practices and skill needs using data on jobs from the public utilities industry (Cappelli and Rogovsky 1993). The workers were asked about the skills they needed to improve performance in their jobs and also about the extent to which they used work practices associated with high-performance systems. The overall results suggest that there are some, although not many, significant differences in skill needs associated with high-performance work. And some of the differences suggested that skill needs were actually lower where there was more high-performance work. For example, skill needs were lower where certain team processes were in place, perhaps because individual workers must function on their own and make more decisions by themselves. As a result, each worker would need more knowledge and skill to perform a given task than when that task is performed in a team where knowledge and skill can be pooled across team members. Overall, the skills that tended to be associated with these new work practices are behavioral skills such as working in teams.

These results suggest that while new work practices may make new demands on worker skills, the demands may not be overwhelming, and they may focus more on behavioral skills than on traditional vocational skills. Thinking specifically about "lean-production" systems in manufacturing, the fact that Japanese auto companies can take inexperienced workers in the United States and in the United Kingdom and produce autos more efficiently than can German companies in Germany where craft work skills are thought to be much higher suggests that the skills required by lean production in particular can be taught relatively easily. New production systems may require learning about concepts such as continuous improvement and statistical process control, but much of the training in Japanese auto companies, in particular, is with these behavioral skills and socialization.

Two other developments related to these trends in work organization are changes in the organizational structure of establishments. The organizational chart that represents the hierarchy inside organizations is getting flatter as the "middle" positions are cut back. The empowerment and team work trends noted above help reduce the need for supervisors, an effect that spills over to higher management (i.e., fewer managers are needed to direct supervisors). New information and control systems automate the compliance functions typically directed by middle managers. And the move toward decentralization—e.g., profit-centered operations—reduces the importance of compliance. Flatter organizational charts mean shorter job and promotion ladders inside the organization. The positions that remain, in turn, become broader.

An overall summary of how work may be changing includes the following conclusions:

- Work practices are changing, with more establishments using teams, employee participation, and other such arrangements. But these arrangements are by no means in all industries and occupations and are not yet close to being a majority. While the prospects for increased diffusion look good, there are also important reasons for believing that there will be limits to the spread of these practices.
- Where new work practices have been introduced, skills appear to be higher, although how much higher is hard to gauge, and the skill demands that have increased seem to focus on behavioral skills.
- With respect to the nature of these new skills, new production techniques like lean production change jobs by broadening them, eliminating certain narrow jobs, and loading their tasks onto others. Teams, employee participation, and the other more popular new work practices often lead workers to move across a much wider variety of tasks that often include supervisory tasks. Behavioral skills and work-based skills in general appear to have become much more important.

- Many of the above changes make it increasingly difficult to use simple occupational titles as a way of identifying the tasks that workers perform. The tasks that a given worker performs are now much broader and more likely to overlap with what workers do. To the extent that workers do have a core set of unique tasks, those tasks may now take up a much smaller proportion of their working time.

Together, the arguments above suggest that there are important changes in skill needs, although they may be less than revolutionary. More attention to measuring workplace skill needs

seems to be in order, particularly as they stack up against the skill set that workers bring to their jobs. The fact that job titles may no longer be good proxies for what one does in a particular job argues for direct measures of tasks performed in each workplace setting. Finally, data collection efforts need to pay more attention to behavioral skills as they seem to be increasingly important in the workplace.

CHANGES IN THE EMPLOYMENT RELATIONSHIP

The second, related work force development is a breakdown in the traditional relationship between employer and employee. The declining obligations and commitments that employers have, especially for their white-collar workers—and the reciprocal decline in the commitments of employees—raise some profound questions about how work-based skills in particular will be developed in the future. This development is closely related to the issue of lifelong learning, that is how the need for skills will be met once workers are in the labor force.[5]

The circumstances that helped create formal arrangements for managing employees in large firms, often referred to as internal labor markets, are changing. Internalized employment arrangements that buffered jobs from market pressures are giving way to arrangements that rely much more heavily on outside market forces to manage employees. There are a number of reasons for that transformation. They include increased competitive pressures on costs and from investors, especially institutional investors, who are demanding higher profits from publicly held enterprises. In addition to the pressures on costs, another factor associated with changing product markets is the need to react quickly to changing consumer demand. The flexibility required to adapt to changing product markets means that fixed costs, including the fixed costs of internalized training and employment systems, become more difficult to support financially. Public policy also contributes to the breakdown of traditional employment relationships. As the legislative protections on regular employees rise, the administrative costs of using such employees rise as well, especially as compared to using contract workers or temporary employees.

Perhaps the most compelling evidence of the changing employment relationship is the decline in job security. One aspect of this change is the continuing pace of downsizing, which appears to actually have increased through the 1990s even as the economy improves. Econometric evidence suggests that the displacement rate for prime age men (35–55) has doubled in the 1990s as compared to the 1970s (Medoff 1993). Employee tenure with their employers' also appears to have declined, especially for older, white men, the demographic group traditionally most protected by internal labor markets. Most important for the discussion here, attachment to one's occupation is actually increasing even while tenure with one's employer is declining (Rose 1995).

The fact that people are staying in the same occupation longer means that there is a greater incentive for them to invest in occupational training because there is a longer time period in which it can pay off. Yet the fact that tenure is declining implies that there is less incentive for employers to provide that training because the contribution from the employee will be made over a shorter period.

The evidence on changes in training is mixed. There is considerable evidence that new work systems demand new and different skills from employees and that employers who are introducing

those systems must train employees to function in them (Osterman 1995). And there is some evidence that this type of training—to improve one's job skills in one's current job— is provided to more workers now than in the past (although the intensity of training appears no greater). But training to learn new jobs has declined compared to earlier periods (Constantine and Neumark 1994).

Many other changes suggest how the attachment between employers and employees may be weakening. The use of temporary employees, for example, has increased by a factor of three since 1985. Even wages exhibit the changing relationship. The returns in the form of higher wages associated with longer service with the same employer have declined sharply over the past decade. Conversely, the costs of changing jobs has virtually disappeared. In the 1980s, for example, workers who changed jobs every other year saw almost the same earnings rise in the late 1980s as did those who kept the same job for 10 years (Marcotte 1994). Several studies report that the pay practices inside firms are now much more subject to market forces than in the past. One particularly striking aspect of that change has occurred with respect to pensions and retirement benefits. In 1979, 83 percent of all the workers who had pensions had defined benefit plans where the benefits were guaranteed and the employer took the risks associated with funding them. By 1988, the most recent data available, finds that figure falling to 66 percent. The change has been due to the growth of defined contribution plans like 401(k)s where benefits are no longer guaranteed and the employees take the risk of maintaining their benefits (Ippolito 1995). Further, with no vesting requirements and no fixed pension costs, these new arrangements create no incentives on either employees or employers to stay together.

The breakdown of attachment between employer and employee raises a number of issues that, in turn, have implications for data collection. Perhaps the most important is the question of how skills and training will be acquired. If workers move between employers more frequently, then the ability of employers to fund training for these workers decreases, at least relative to the demand. Workers are increasingly expected to manage their own careers and seek out training themselves to improve their skills. Especially if workers are staying in the same occupations longer, they are more able to reap the gains of improved skills. We should expect much more of a market to develop for training as workers look outside their current employers for training.

As workers move from employer to employer, we might expect them to stop at schools in between to upgrade their skills. Here the notion of lifelong learning has some powerful policy relevance as the demands on schools will change. In terms of data needs, it is important to learn what these returning workers will demand from schools by way of upgrading their skills; for example, what kind of work experiences create what skill needs at which point in one's career? What makes some workers come back to postsecondary institutions while others go to vendors or alternative providers?

Markets require information. In this case, the labor market will require more information on the skills that workers have as they change jobs, and employees will want to know both what skills are required in different settings and where they can go to get those skills. We might expect greater data needs both from and for all three groups—employees, employers, and schools.

One way to think about this new situation is that it may repeat the school-to-work transition problem several times over a worker's career. All the issues about how to make learning more

responsive to workplace needs, how to signal skills to employers when leaving school, and so on, get compounded when one is going back-and-forth from school and work.

FUTURE DATA ISSUES

The developments outlined above serve as background to some long-standing questions for which additional education and work-related data are needed. These questions are organized into two major headings:

What Does Work Demand from Employees?

What knowledge, skills, and abilities (KSAs) are required by people entering the work force or already in it that could be met by the educational system, broadly defined? This seems like a unnecessarily general question, but it helps to set up the choices that must be made by policymakers in defining data collection and research questions that can be tracked more easily.

Perhaps the first choice is what does it mean to say that work "requires" something from employees? Does that mean, for example, the requirements needed to get a job—the type of KSAs typically found in job descriptions like those in the *Dictionary of Occupational Titles*? Such requirements can be thought of as either the minimum needed to carry out a job or to be competent at it. Or does it mean the KSAs "required" to excel in a job, associated with improved job performance? The two may be very different and not necessarily be matters of degree. Excelling at a job, for example, often means finding ways to go beyond the current standards as defined by job descriptions or finding ways to alter the task requirements.

The minimum competency approach is not really an empirical research question in the usual sense. It is not, for example, derived from the actual experience of employees. Rather, it is more a deductive process based on the a priori requirements as articulated by industrial engineers who design the jobs. Job analyses in personnel psychology essentially collect this kind of information. The analysts ask either experts or sometimes the employees themselves to identify the tasks that they perform and then use various taxonomies to organize the requirements into KSAs. Some of the taxonomies are organized around the traits that employees need to do the jobs, while others are organized around the characteristics of the tasks themselves. The skills generated by the SCANS Commission are based on job analyses that mix the trait and task approach.

Most of the research on whether skill requirements are changing have been based on job analysis-type data like that contained in the *Dictionary of Occupational Titles*. It is important to understand what exactly such measures can tell us. They capture a point-in-time assessment of what employers ask employees to do with respect to the organization of work. They do not attempt to assess whether what they are doing makes sense and whether it in fact contributes to performance. For example, a job analysis of manufacturing jobs 10 years ago would reveal a set of required KSAs (e.g., emphasizing compliance and downplaying initiative) that now are seen as retarding improved

performance in the light of "high-performance" work organization in manufacturing that is both dramatically different and apparently much more efficient than in the past.

Job analysis data might therefore not be especially valid as an indicator of what skills are really needed in the future. What employers are doing at any point may not be optimal and in any case is always likely to change. (Many observers suggest that we have a skills problem in the United States precisely because we set out expectations for the educational system based on what employers demanded from front-line workers 10 years ago, which was very little.) Job analysis data over time might be a better indication about the trends on how employer requirements have changed.

A related use of job analysis-style information is to estimate how changes in the distribution of employment across occupations may affect future skill demands in the economy as a whole. For example, a shift in employment from manufacturing toward clerical jobs means that the skills required in the average job will change. But the problem noted above still applies: Current skill requirements of jobs may not reflect optimal or even future requirements.

Validating job analysis data is problematic without some other independent set of information on job requirements. More to the point, requirements from job analysis data are rarely related to actual job performance measures. Again, job analysis data indicate only what is required for minimum performance and do not suggest what KSAs are required for superior performance. It could well be that the KSAs required for superior performance in a job are very different from those described by job analyses for minimum competence. The way to tell, of course, would be to examine the relationship between KSAs and actual job performance. Such relationships answer a different question—what predicts better performance? The ontology behind this approach is very different than that described above. While job analysis is a kind of deductive process where a given task is mapped onto KSAs using a set of established algorithms to identify job requirements, real validation efforts reveal underlying relationships between KSAs and performance by looking for statistical relationships. There is no reason to expect that the two approaches will yield the same results.

The validation approach of comparing actual job performance to worker characteristics has several important advantages as a means for identifying the KSAs that are important for work. First, it does not require algorithms or judgments about linking tasks to KSAs. Nor does it require mapping out what an individual employee actually does on the job. As noted earlier, identifying the full range of skills one performs on the job becomes increasingly difficult as jobs become broader, and more flexibly defined, and workers are given substantial autonomy over both what tasks they perform and how those tasks are carried out. As noted earlier, what an individual actually does in a particular job title may well vary day-by-day now as well as by situation (e.g., two secretaries with the same job title may do very different things depending on who their boss is).

Further, the validation approach of looking at actual performance makes it much easier to see relationships with educational characteristics. With job analyses, the particular set of KSAs being labeled varies with the type of job analysis chosen. And mapping a given taxonomy of KSAs onto educational characteristics is not at all straightforward. For example, if a job analysis reports that a given job requires a high level of problem-solving skills, what does that say about educational

requirements? Does it mean that graduates will do better with more math or logic courses, or is the problem-solving so contextually oriented that something like engineering courses are really what is required? The validation approach would provide direct answers to these questions by showing the effect of different course-taking patterns on student performance.

Job Analysis Data

The National Job Analysis Study currently being undertaken by American College Testing represents what will be the best information available on current job requirements for the economy as a whole. It is designed to provide something like minimum competencies for broad clusters of jobs across the economy as a whole. In terms of additional data collection in this area, the most useful approach would be to repeat something like this study at a later date in order to assess whether these average competencies are changing—not only whether employment shifts across occupations are affecting average skill levels but also whether the skills of particular occupations are changing.

Beyond the job analysis-style assessment of average competencies, which are essentially impossible to validate, it is less obvious how this job analysis data can be used. It will represent something like a taxonomy of relevant skills that has been grounded in field-based experience. Not all of the skills it identifies will be relevant for education, however, as some may be quite job- or context-specific. Most observers would agree that the focus for education should be on the KSAs that are at least to some extent cross-functional, extending beyond individual jobs and, at a minimum, onto careers within general occupational areas. Determining how many KSAs are truly relevant across all jobs is a difficult question, and whether policy makers want to focus down to the level of specific occupations, losing generality in the process (as the National Skills Standards Board is doing), or aggregate up to some higher level, thus losing specificity, is a difficult choice.

The skill information from the National Job Analysis Study can also be used as a taxonomy for collecting further information on job requirements. For example, if it turns out that certain skills feature prominently across occupations in the job analysis data, then perhaps we need to collect data on those skills—e.g., how widespread they are—for other analyses.

The first issue might then be which skills to include. The distinctions used in the *Dictionary of Occupational Titles* between basic, cross-functional, and occupation-specific skills seem to be the most appealing criteria to use as a way of including skills into a classification scheme. They strike a reasonable trade-off between parsimony and richness and get at the kind of information that is relevant in the labor market. Campbell (1994) offers a good assessment of what is required to make such an arrangement work.

But collecting data on the KSAs relevant for education is a problem. Stevens (1994) and others have raised the important practical issue of the limits imposed on any classification system when it goes into the field. The issue of parsimony needs to be considered from the perspective of the NCES operations that are compiling the data. For the reasons noted above, it is unlikely that simply asking a respondent's occupation will provide accurate information about what he or she does

on the job and what skills are needed. Many more detailed questions are required, but a population survey has a fixed and relatively small number of questions it can ask.

Consider the current arrangements at the Census, for example. The Current Population Survey (CPS) asks respondents about their business or industry, the kind of work they do, and their most important activities at work (Census 1989). This is not a great deal of descriptive information about the job. Classification clerks then take these responses and aggregate them into occupational codes. In about half the cases, employees believe that their occupation is something different than does their employer (Mellow and Snider 1995). At least half the time, then, one of the parties—employer or employee—is wrong in labeling an occupation.

In other data collection efforts, respondents give the interviewer their job title. Dempsey (1993) suggests that about 10 percent of employers participating in the Department of Labor's Occupational Employment Survey simply submit their current job titles for Census data collection efforts. Researchers then use information from the D.O.T. or other sources to infer information about what skills are required for that job title, ultimately generating estimates for the sample about skill requirements and other issues. The problem, of course, is that the job title the respondent has in his or her organization may be idiosyncratic. It may not correspond well at all to the title that someone in another organization doing the same tasks may have. As noted above, organizations may be getting more idiosyncratic in their job titles, making it even less desirable to let respondents classify themselves.

Interviewers really need to ask respondents directly about their jobs in order to get detailed information on tasks and skills. The experience in Ohio suggests some lessons for how a data collection system might be implemented. Somers (1993) reports that the Ohio Bureau of Employment Services resorted to a series of keywords and computerized text searches for matching workers with jobs, adopting aspects of the Canadian JOBSCAN system for mapping work-related skills that rely on simplified checklists, like keywords, which can be updated easily as jobs change. Perhaps it is possible to use simplified taxonomies like these for measuring the skills required in jobs.

It is important to remember, however, that all of this information is still only about *jobs*. It reflects only minimum requirements of the kind described earlier and cannot be used for any validation efforts relating skills and performance. That requires collecting data on the KSAs individuals possess and then comparing them to some measure of actual job performance.

What Predicts Workplace Success?

As noted above, job analysis-style information that establishes minimum competencies is not the same thing as identifying success on the job. Efforts to identify the characteristics of workers that predict labor market success, almost uniformly defined as wages by labor economists (sometimes unemployment or other labor force status measures are used as well), explain relatively little of the total variance in the outcome or success measure; in fact, they explain rarely more than about a third. Personnel psychologists generally use broader, but potentially more subjective, measures of job performance such as the evaluations of supervisors. Their efforts at predicting performance are more

successful, sometimes explaining as much as half of the variance in outcomes, but the studies have other methodological drawbacks such as non-random selection.

One of the most basic needs for research is simply to provide some validation on the basic issue of what work demands from employees in terms of KSAs by relating those KSAs to actual job performance. Once we have job analysis-style data, can we show that those KSAs in fact predict an individual's job performance? That need, in turn, makes some important demands on data. The first, as noted above, is simply to measure the relevant KSAs in employees. This demand leads to an important question: What is the boundary between KSAs obtained from education and from other areas?

The KSAs that are presumably of greatest interest to NCES are those that are related to educational institutions, those that one would expect to be learned in schools. But in practice, the KSAs relevant to success in the workplace are likely to be learned in the family, in school, and in a wide variety of settings that are difficult to separate. This is especially the case where school-to-work programs have been introduced with the goal of blurring the distinctions between these categories of learning.

One approach to this problem is simply ignore it, and to rely instead on traditional measures of academic achievement that measure classroom learning. School-based credentials like degrees, grades completed, and achievement test scores measure what has been presented to students in the school setting. No doubt they are unlikely to represent all or perhaps even most of what is relevant to workplace success. But when related to measures of such success, they do allow one to address whether education matters for workplace success and, if so, which aspects matter. This is obviously more limited than knowing what workplace success demands in terms of KSAs. But knowing how traditional academic achievement matters for workplace success would still be a considerable achievement over where we are now.

Within the general heading of understanding how educational experiences affect employment outcomes are three subquestions:

Better Data on KSAs

Perhaps the first question is simply to develop a better understanding as to what education-related characteristics, or KSAs, determine how well a student does in the labor market. The place to start is to get better information on what the components of an individual's KSAs might be. As noted earlier, traditional measures of academic achievement help us understand how student achievement in the context of current curricula and pedagogy affect labor market success. But this is still a bit of a black box in that we cannot unbundle the subcomponents of academic achievement. For example, if grade point averages predict job success, is the power of the grades coming from the academic knowledge they measure, the comportment aspects they capture (attendance, perseverance, and so on), or the more general problem-solving and organization skills that help determine academic success?

Within the context of academic success, we first need better measures of academic achievement that go beyond traditional grade point averages. The data sets that include standardized

test scores are clearly an improvement over grades alone in that they allow us to measure cognitive performance independent from the classroom experiences that affect grades (attitudes, participation, and so on). Several NCES data sets already include such measures. Including more general cognitive ability tests like the General Abilities Test Battery (GATBy) in data also captures something different from subject-based achievement tests. These measures have contributed in important ways to research on labor market outcomes (Tyler, Murnane, and Levy 1995). One problem with such tests, however, is that they tend to be unreliable unless students have a real stake in doing well on them; tests that are administered simply for the purposes of the survey will find students not making the effort to do well on them, thus biasing the results. It is not obvious how to address that problem, which means that samples using such tests will have important biases (either they exclude those who do not take them, a group that is systematically different in other ways, or they include them and somehow try to account for the fact that their performance will be worse).

Currently, one of the most fundamental questions in the topic of employment is the extent to which job performance is driven mainly by cognitive ability, as some have argued (Ree 1994). If this is so, then perhaps curriculum and pedagogy should be redesigned to emphasize cognitive development. But we need better data and more research to identify whether this really is the case. For example, the data used to argue for the importance of cognitive ability in personnel psychology typically do not include measures of an individual's educational experiences; therefore, it is impossible to tell whether the measures of cognitive ability in fact stand as proxies for aspects of education that covary with cognitive ability.

It is also clear, however, that a wide range of important educational experiences are not examined by current data. Extracurricular activities, for example, appear in the research noted above to be very important in shaping workplace performance but are not typically measured in any detail in current surveys. Particularly with regard to the transition from school-to-work, some of the most important experiences facilitating that transition may take place outside of school. And while basic information on work experiences is currently collected in several NCES databases, it would be helpful to have more detailed information on what actually happens to student workers in the workplace. For example, how are they supervised? Do they receive any formal or informal training and, if so, of what kind? What is the nature of the tasks that they perform? Questions like these are very important in understanding what helps students make the transition to the workplace and in designing curricula to facilitate that transition (see below).

More generally, work-based skills and competencies are not directly measured by any of the national probability datasets, nor are behavioral skills or dispositional characteristics like personality that both prior research and commentary suggest are crucial to job success.

The term “behavioral skills” is a code word for a range of knowledge about issues such as group and individual behavior, interpersonal and self-management skills, and attributes and abilities. The first problem with collecting data on behavioral skills, indeed on any work-based skills, is how to measure them. There are a number of competing taxonomies for such skills like the trait-based job analyses in personnel or the SCANS skills used in public policy. Every taxonomy “cuts” the KSAs in a slightly different way.

The problem for NCES in collecting data on work-based and behavioral skills is first to choose a taxonomy for measuring those skills. The key issue is to choose a taxonomy that does not leave anything out and that avoids lumping important concepts together. The SCANS skills, for example, seem to put together many distinct behavioral skills into the same categories (e.g., self-management and interpersonal skills), making it difficult to interpret relationships with those measures. It might also be important to anticipate which of the various taxonomies will come to be accepted in future policy discussions. Will American College Testing's National Job Analysis Study, for example, be embraced by the research and policy communities, and should NCES use its taxonomy of skills for collecting data on work-based skills? One sure bet is that no single taxonomy will be embraced by the research community. There have been decades of debate and contention regarding the appropriate methods for doing job analyses with no clear consensus emerging as to the "best" taxonomy, because each represents trade-offs on issues about which reasonable people can and do differ.

Perhaps the best advice on this issue is to have the various government agencies interested in measuring work-based skills agree on a taxonomy and get on with it. Objections will be raised no matter what is chosen, but if there is agreement among the government players, the taxonomy selected will become the standard: "If you collect it, they will use it."

How to measure work-based skills, particularly behavioral skills, is a more complicated problem. It may be possible to proxy skills with certain credentials like coursework related to behavioral skills. While taking a course in interpersonal skills may not seem like a good proxy—indeed, it may simply select in those people who have bad skills and are taking it because they really need help—the same procedure is generally used to measure one's academic skill base in a subject area like math. In the absence of clear credentials, it becomes difficult to rely on self-reporting, and surveys must find some other way to measure skills. In the area of academic achievement, a series of well-established standardized tests are available for measuring subject knowledge and various abilities. There are no real equivalents yet on the behavioral side, although there are well-accepted tests in specialized areas like personality profiles. But someone will certainly seize the enormous opportunity that tests of behavioral skills offer in improving employee selection, and those will soon be available.

Better Measures of Education Institutions

If we had a better understanding of which student characteristics lead to success in the workplace, it would then be important to learn what characteristics of educational experiences, broadly defined, help produce those characteristics.

The "toe-in-the-water" approach to additional data in this area is to collect further, more detailed data on classroom experiences. Most of the research on education and labor market outcomes has been limited to looking at gross measures of educational attainment—years of education completed and degrees conferred. Perhaps the most important innovation in contemporary research has been to add detail to those existing measures. The NCES data on student transcripts, for example, has made possible new research on the effects of patterns of course taking on labor market outcomes (Altonji 1995). This research has been well received and has already contributed in a central way to policy debates such as the relative returns to attending 2-year versus 4-year

institutions (Kane and Rouse 1995). What is perhaps most surprising about this line of research is how long it has taken to get started and how much remains to be done. It is possible to count almost on one hand the number of studies that have looked at the content of student coursework as it affects labor market outcomes.

A few studies in personnel psychology have explored the impact on job performance of student experiences in addition to course-taking patterns. These include, for example, studies of extracurricular activities where the results suggest that these experiences are very powerful predictors of job performance, more powerful, in fact, than academic performance (Bray, Campbell, and Grant 1974).

A related development, also in its infancy, has been to look at the characteristics of educational institutions as organizations that affect the labor market performance of their graduates/attendees. There are many studies that look at how the characteristics of postsecondary schools and teachers affect the academic achievement of their students (see Hanasheck et al. 1994 for a recent review), but again, very few that link those characteristics to labor market outcomes. For example, no studies have looked at the relationships between aspects of how schools are organized and the labor market performance of their students (Johnson and Summers [1993] review this literature at length.)

Among the very few studies that attempt to link school characteristics to labor market outcomes of their students are Crawford, Summers, and Johnson (1994) for secondary schools, and Daniel et al. (1995) for higher education institutions. The results suggest that the characteristics of these institutions do matter, but the measures are aggregated at a level that makes it difficult to see relationships with specific practices and to offer detailed guidance on organizing schools.

The data problems in linking school characteristics and labor market outcomes begin with the fact that most of the surveys that collect longitudinal labor market data are national probability samples where it is unlikely that many respondents will come from the same institutions. The pathbreaking analyses will be to look within institutions to see how variations in education experiences affect student performance—both traditional academic achievement and labor market outcomes. To illustrate, data that might find better student performance associated with attending small liberal arts colleges is confounded: Is the better performance the result of smaller class size, small academic communities, the typical liberal arts curriculum, or the characteristics of students selected into such schools? We would need to look at the variance in experiences within these schools in order to answer those questions.

The data required to address these within-institution questions are considerable: first, the data must be longitudinal, following students through their postsecondary experiences and into the labor market; second, they must represent samples of reasonable size within postsecondary institutions; and third, they must include a wide range of such institutions. These data needs are considered in more detail below. If available, they would offer an enormous research opportunity for relating traditional measures of academic achievement and school characteristics to labor market outcomes.

Better Data on Work Outcomes

The arguments above suggest the need for better information about the knowledge, skills, and abilities that individuals possess in order to explain work outcomes and, in turn, determine what KSAs are really demanded in the workplace. Even with this better information, however, there is a weak link in the analysis, and that is the measure of workplace outcomes and performance.

As noted earlier, the majority of studies relating education and work outcomes use wages as the measure of "success" on the grounds that superior performers will be rewarded with higher wages, other things equal. But there are some obvious difficulties with that approach. For example, wages are driven perhaps most strongly by occupational choices and not performance within an occupation; the best school teacher in the world still earns less than a mediocre investment banker. Occupations differ greatly in how wages relate to performance. A good sales associate may earn substantially more than a poor one, but a good teacher is likely to earn about as much as an average teacher. In general, the relationship between performance and compensation may not be especially strong across the economy.

It is certainly possible with modern econometric techniques to address some of these issues. For example, looking at wages within occupations, controlling for employment status (i.e., wages conditional on having a job and on working hours) and other factors that might affect pay, may address some of these issues. But short of perfect modeling, these are at best imperfect adjustments. For example, someone who pursues his or her occupation in the non-profit sector of the economy will earn less. The characteristics that lead someone to make that decision (e.g., attending a college with public service requirements) will turn up in a validation exercise as being negatively associated with earnings and, in turn, appear as something that actually hinders workplace performance.

Some improvement comes with expanding the range of labor market data on individuals to include, for example, spells of employment, long-term career earnings, training received and career mobility, job and life satisfaction, and so on. Ultimately, however, we need better information about the nature of work performance for individual workers.

Specifically, it would be important to know not only whether a worker is doing well or not but also which aspects of their performance are good and which are poor. Ideally, we would like that information in ways that tie directly into KSAs—are there skills that the employee seems to lack, for example, that are associated with poor performance? Such information would be especially helpful to know for new entrants/school leavers where the link between education and performance may be most clear. There is a perception, for example, that the school-to-work transition problem is in part due to comportment problems and poor self-management skills among school leavers. Detailed information on their performance would be especially useful to address that issue.

A survey conducted by the National Foundation of Independent Businesses (NFIB) offers one example of alternative performance data on employees. The survey of employers asked a series of detailed questions about the last employee hired and his or her job performance (actual versus

expected). The Department of Labor in the State of New Hampshire collected similar data on school leavers by going to their employers and asking detailed questions about how those individuals were performing in the workplace. Personnel psychologists routinely collect such data on a wide range of performance outcomes, including promotion potential, organizational citizenship, and so on.

The main difficulty with alternative performance data is in collecting it. Unlike wage data, these data cannot be self-reported accurately, and many questions must be used to produce reliable scales for each concept. Such data must be collected from employers. Surveys like the General Social Survey and the National Organizations Survey have collected matching data from employers and their employees by asking respondents to identify their employer and then contacting and surveying the employer. The additional problem with individual performance measures is that it is unlikely that a centralized personnel office could complete surveys about aspects of a specific employee's performance, especially in large establishments. Supervisors within the establishment may have to be enlisted to answer the questions, raising rater reliability issues and reducing the expected response rates. When personnel psychologists collect such data, it is typically within a single organization where the organization's own performance measures can be used. These may be consistent within that organization, but they are unlikely to be consistent across different organizations.

Work Performance Beyond the Individual

As noted earlier, the interest in how education affects workplace performance has been driven not just by the belief that it might improve an individual's performance and earnings but also by the view that it might make both establishments and economies more productive and effective. Research such as that performed with the National Employer Survey (EQW 1995), which finds that establishments with a more educated workforce, other things being equal, are more productive, has been the focus of considerable policy interest.

How NCES might develop data to expand the measurement of performance is worth considering. The first issue to confront is that it would require performance-based information on groups larger than individuals—teams or work groups, establishments, and so on—an effort that might seem far beyond the traditional paradigms of NCES data collection. But there are some exceptions even with the data that NCES already collects. For example, it collects detailed organizational information on one type of operation; schools, using the Schools and Staffing Survey. Studies examining how the educational background of school staff affects student performance are already relatively common. It would not require much new data on the educational experiences of teachers and administrators to examine the relationships between establishment-level performance and the particular experiences of school staff.

Beyond this education-specific setting, there may be ways to join forces with other establishment-level surveys in order to examine the performance effects of education.

How Does Work Affect Education?

While most of the recent policy interests seem to be focused on the question raised earlier of how education can contribute to workplace success, the more traditional and equally important

question is how work experience affects academic achievement. How secondary school work experience affects students' educational performance is a question with a significant research tradition, but several more contemporary issues also demand attention.

Given that so many students work while attending school and the trend toward combining work and school in postsecondary education seems to be increasing, it is very important to know how traditional work experiences (i.e., part-time jobs) affect educational performance. We need to go well beyond existing research, which has focused mainly on how hours of work affect student classroom achievement, to understand how the characteristics of that work experience affect academic performance. The general public understands that the nature of the work experience is crucial to educational success, as evidenced by the different language we use to describe different student work experiences (i.e., internships versus part-time or summer jobs). Consider some of the following research questions:

- Especially for secondary school students, what effect does working in a stereotypical fast-food or low-skill job have on academic performance? When, for example, student workers are often supervised by school dropouts barely older than the students themselves, are there negative "modeling" effects that lead to worse academic achievement?

- Especially for postsecondary education, does having a "better" job that offers more opportunities for learning and advancement while attending school actually contribute to dropping out as employers pull the best students out of school and into full-time jobs? Or does it allow more students to complete school by increasing their resources? Does it change their course-taking patterns and choice of major? Do students with more work experience have a smoother transition to the workplace after graduation?

- What effect does work experience have on KSAs other than the classroom-based knowledge measured by traditional achievement tests? Do different kinds of work experience provide alternative vehicles for learning SCANS-type skills, for example?

- How do different kinds of work experience affect postsecondary school experiences—attendance, completion rates, course and major selection, and so on?

- Finally, how does work experience shape the demand for continuing education? Do different kinds of work experience make it more likely to pursue postsecondary education? For example, does a part-time job in a hospital, where one learns about all kinds of careers that require further training, make one more likely to pursue further education than if one did the same kind of unskilled work (e.g., janitor) in a different setting? Even for students who do not attend traditional postsecondary institutions, do different kinds of work experience make them more likely to pursue skills and training through other avenues?

In Secondary School

Researchers have argued back and forth about the effect on student achievement and subsequent educational plans of working while in school. With few exceptions, this research has focused on the quantity of work, with relatively little attention paid to the quality of the work experience. As argued above, better information on the characteristics of a student's working experience would help considerably in understanding the real impact on education. Such information and data are a special priority at present given the introduction of school-to-work transition programs across the country and the need to understand what makes them successful.

The type of evaluation of vocational education programs recently conducted as part of the legislative reauthorization would also be enhanced considerably by knowing the characteristics of the work experience in those programs. It might well be, for example, that there are no real differences between youth apprenticeship programs and cooperative education programs and that the apparent variance in their results is simply due to the characteristics of the work experience in each setting.

In Postsecondary School

All of the above issues apply to student experiences in postsecondary school as well, although they have been far less researched. Student working hours and experiences may have important impacts on academic performance as well as various kinds of institutional arrangements such as co-ops programs and summer internships. Whether and how much students work in school is linked closely to issues of student financial aid and school resources, another important policy issue.

Lifelong Learning

The issue of education after entering the labor force needs to be put squarely on the research agenda. As the length of time many students attend postsecondary school gets longer and increasingly is combined with full-time employment, it no longer makes sense to think of this as simply delays in graduation. It may be more appropriate to think of this situation not as a transition period to graduation but as a new and stable pattern: going back and forth from work to school, taking new courses as workplace demands require them, and possibly making career and work changes as new skills are acquired. All of the above issues as to how work experiences shape educational choices and outcomes apply to these new “lifelong learners” as well.

RECOMMENDATIONS FOR DATA

Most of us would be delighted to see NCES develop new data sets specifically tailored to meet some of the concerns noted above, but given the tremendous investments required for such efforts, it would be impractical at best in the current climate of fiscal restraint to make such recommendations. In fact, some of the important questions can be addressed using existing data, and

relatively simple additions to the data series currently maintained by NCES would address many of the remaining data needs.

The most basic data need is to have information in the same data set about an individual's educational and work experiences. An issue that is integral to many of the more specific questions raised above is simply to get a better understanding of what demographers refer to as the "life course" of young people. Have the paths from school-to-work or secondary to postsecondary school changed? Consider some of the basic factual questions embedded within that more general question for which we currently do not have good answers:

- Are more postsecondary students working full time?
- Has the pattern of "articulation" or transfer of students from less-than-4-year to 4-year institutions changed?
- Are postsecondary school graduates returning to school after entering the work force to upgrade their skills?
- How many secondary school students participate in school-to-work programs?

NCES already maintains a number of data series on individuals and their educational experiences, as shown in Figure 1.

Figure 1—Availability of data in NCES sources that can be used to measure components of school-to-work

Type of data & school-to-work components	HS&B	NELS	NPSAS	BPS	B&B	SASS
Type of data source	longitudinal	longitudinal	cross-sectional	longitudinal	longitudinal	cross-sectional
Years of collection	1980–86, 1980–92	1988–94	1987, 1990, 1993	1990–94	1993–95	1988–91, 1994
Level at which data are specified	student & school	student & school	student & institution	student & institution	student & institution	school
School-to-work components						
A. Educational preparation for work						
1) Educational Attainment	yes	yes	yes	yes	yes	NA
2) Postsecondary enrollment & persistence in school	yes yes	yes yes	yes short-term only	yes yes	yes yes	NA NA
3) Transcript data	postsec. only, sec. & postsec.	secondary	no	no	forthcoming (postsec. only)	NA
which can be used to:						
● distinguish among students with similar degrees	yes	yes	no	no	forthcoming	NA
● measure attainment	yes	yes	no	no	forthcoming	NA
● specify a career major	yes	secondary only	no	no	postsec. only	NA
● assess exposure to all aspects of an industry	yes	yes	no	no	postsec. only	NA
4) Grades	yes	yes	student report	student report	forthcoming	NA
test battery scores	yes	yes	no	no	no	NA
which can be used to:						
● develop gain scores	yes	yes	no	no	no	NA

Figure 1—Availability of data in NCES sources that can be used to measure components of school-to-work—Continued

Type of data & school-to-work components	HS&B	NELS	NPSAS	BPS	B&B	SASS
B. Work experience						
General availability of measures						
Employment Status	monthly 1980–86, 1982–92	monthly 1992–94	annualized	monthly 1990–94	monthly 1993–94	NA NA
Wages	1980–86 only	yes	limited	yes	yes	NA
Earnings	yes	yes	yes	yes	yes	NA
Avg. hours per week	1980–86, 1982–86	yes	yes	yes	yes	NA
Occupation	yes	yes	no	yes	yes	NA
Industry	yes	yes	no	yes	yes	NA
Relatedness of employment to education[2]	student report & linked codes	student report & linked codes	student report & linked codes	student report & linked codes	student report & linked codes	NA
Availability of measures by topic						
1) Employment experiences in high school	yes	yes	no	no	no	NA
2) Employment exp. in postsec. enrollment	yes	yes	limited	yes	yes	NA
3) Employment exp. as an outcome	yes	yes	no	yes	yes	NA
C. Patterns & processes of articulation						
1) Secondary to postsecondary	yes, with 10 yrs. post HS	yes, with 2 yrs. post HS	no	no	no	NA
2) Postsecondary to postsecondary	yes	no	no	yes	perhaps w/transcripts	NA

Figure 1—Availability of data in NCES sources that can be used to measure components of school-to-work—Continued

Type of data & school-to-work components	HS&B	NELS	NPSAS	BPS	B&B	SASS
3) HS or postsec. to employ.	yes	yes	no	yes	yes	NA
D. Availability of institutional resources						
1) Number of HS w/work prep. programs	yes, but dated	% of students in programs	no	no	no	yes
2) Number of postsec.inst. w/work prep. programs	no	no	yes	yes	yes (BA/BS only)	no
3) Availability of teachers to teach integrated academic & applied curricula	no	perhaps, but not representative	no	no	no	yes
Background items Student characteristics	yes	yes	yes	yes	yes	no
Family characteristics	yes	yes	yes	yes	yes	no
School or institutional characteristics	yes	yes	yes	yes	yes	yes
Community type	yes	yes	no	no	no	yes
Attitudes and expectations	yes	yes	some	some	some	no
Population characteristics	yes	yes	yes	yes	yes	yes

NOTES: "Yes" indicates that the data set includes items in which the school-to-work element can be measured; "No" indicates that the database does not contain such items; and "NA" means not available. Other entries indicate that the topic is covered by items in the data set, but that coverage is limited as described.

SOURCE: Medrich, E. and Tuma, J. *School-to-Work Data Available in NCES Data Sources*. 1995. Washington, D.C.: National School-to-Work Office.

One can see even from this brief description how rich many of these data sets are in terms of information on education. Several of the data series, like High School and Beyond, the National Educational Longitudinal Study, and the National Longitudinal Study of the Class of 1972 involved collecting data from a respondent's school. Even the richest of these surveys, however, are thin on the following attributes:

- *Content of educational experiences*. Only the three surveys in the above paragraph and the Baccalaureate and Beyond survey have transcript data. And, as noted above, it is difficult to know much about what students actually learned in those courses without more standardized instruments like achievement tests. It would also be helpful to have information on pedagogy—did the classes require written assignments or lab work, was there class discussion or team projects, how big were the classes, were the exams essay or multiple choice? These factors are perhaps even more important to the current debate about education reform than are curriculum issues.

- *Information on relevant KSAs*. None of the NCES data sets currently collects information on behavioral skills or on the kind of work-based skills described by the SCANS report or similar exercises.

- *Details on work experience*. Understanding how work affects education requires knowing about a respondent's work experience. The data currently collected in NCES surveys looks at what might be called outcomes of work—job titles, industry, hours, and wages. What we do not know is what students actually did on the job. What kind of training or supervision did they receive; what tasks did they perform; did they participate in decision making, and so on. As noted earlier, job titles never conveyed much information on these issues, and there are good reasons for believing that they will be even less reliable in the future.

- *Information on job success*. The current NCES data sets have only information about wages and earnings that have limits as proxies for job performance. As noted earlier, it is important to know exactly where workers had success, where they had difficulty, and what skills or tasks were in deficit.

- *Details on employer practices*. If the interest in lifelong learning is real, then it is especially important to know what pushes people back to school after they have joined the work force. The nature of work organization no doubt plays some role in that decision as does a series of employer practices such as tuition reimbursement plans or career planning and progression programs.

Strategies for Collecting New Data

New Data on How Education Affects Work

Clearly the best approach for addressing at least some of the data needs outlined above is to leverage existing data sets by adding data to them. High School and Beyond and the National Longitudinal Survey of 1972 have important attributes in that they contain some reasonably detailed

information on education experiences, and, more important, they contain a long enough time series to identify a respondent's long-term job success. Such information is especially important for assessing the effects of education on work. The drawback to such data, however, is that the respondents have typically been away from formal schooling so long that it is very difficult to collect additional information from them about educational experiences.

HS&B and NLS–72 can be supplemented, however, to address some of the questions noted above about the effects of education on work. First, simple questions on job success could add information to the wage data. For example, a few questions asking about job content, a respondent's position in the hierarchy, and mobility would help identify workplace success. Self-reported data on skill needs would be easy to collect. When related to earlier data on educational experiences, these responses would help identify how work affects job success.

These two data sets in particular would be especially useful in addressing some of the lifelong learning issues noted above. Specifically, what makes an individual seek further education, and if he or she does, what kind of education (topic and provider) does he or she seek? Some of the information on educational choices over a lifetime is already in these data sets. What needs to be added are questions about work experiences. First, what is it about the type of work a respondent performs— tasks and job content as noted above—that pushes them to get further education? Is more challenging work the driver, or is it that those who go back for more education eventually get more challenging jobs? Second, what is it about employer practices that encourages lifelong learning? Is it financial support in the form of tuition reimbursement, or is it incentives like merit-based pay and promotion systems? Together with the job performance information above, these new data would allow researchers to know whether lifelong learning contributes to job performance and, if so, the kind of learning and education experiences that affect job and labor market performance.

Several of the problems noted above hinge on getting data about employers such as performance measures that cannot be obtained from surveys at the individual level. Employer-level data is important for addressing questions such as the following:

- How might different aspects of education in a work force affect organizational performance?
- How does having a more educated work force affect how work is organized or other issues of organizational operations?
- What characteristics of employers (and jobs) contribute to increased use of postsecondary education among employees?
- To what extent are specific postsecondary courses and programs substitutes for firm-provided training, especially at the community college level?
- What are the skills that employers demand from their work force, and how might they be changing?

Such information comes from establishment-based surveys like the EQW/Census National Employer's Survey. But NCES does not maintain establishment-level data sets. Such establishment-

level data would still leave one with the problem of getting detailed information on the educational experiences of individual employees.

The ideal solution is to provide matching data for employers and employees, asking the relevant questions for each group and then putting the two sets of data together. Two approaches for doing so and constructing sampling frames have been used. The first is to survey a probability sample of individuals, asking them about their educational experiences and so on, and to identify their employers. The next step is to go to their employers and survey them about their practices and performance. This technique was used by researchers conducting the National Organizational Survey (NOS) funded by the National Science Foundation. They used questions from the General Social Survey (GSS) of individuals to identify employers, and the GSS data on individuals was then matched to the NOS data on organizations. For NCES, the best method would be to ask the respondents in existing surveys like NLS–72 and HS&B to identify their employers, survey the employers, and then match the data. One problem with this approach, of course, is that there is only one respondent/employee per employer, and it is very difficult to use the experiences and characteristics of that respondent to generalize about the work force as a whole.

The alternative is to conduct a probability survey of establishments and then survey the employees within that establishment. This is the approach currently being used by the Bureau of Labor Statistics in its training surveys. It is an expensive process, as it requires getting information about the work force from each employer (i.e., the sampling frame) and permission to survey their employees. Another approach under consideration by the EQW/Census National Establishment Survey and used by Statistics Canada in their training survey is to try to survey employees in establishments without knowing the sampling frame in each establishment. But even with this technique, the process is expensive and time consuming. NCES does not have to address every data need itself, and establishment-level data are probably not within its comparative advantage.

The questions noted earlier of relating educational practices at the institutional level to student job and labor market outcomes raise very similar problems for data. Addressing such questions requires matching longitudinal data on individuals and their work outcomes to detailed data on the characteristics of their educational institutions. And many of the same problems of matching individual and organizational data appear here as well; specifically, the need to have many observations from the same educational institutions in order to estimate the effects of within-organizational practices.

Here, the best strategies for data collection do not seem to leverage in any obvious way off of existing surveys. One approach might be to develop a targeted sample of institutions whose education practices and arrangements seemed especially noteworthy or representative, and then to follow a representative sample of their graduates over time to examine their labor market performance. One could then use the data to relate practices and experiences at the classroom level, within institutions, to workplace outcomes.

New Data on How Work Affects Education

Understanding how work affects education is an issue that seems especially within the traditional purview of NCES. It requires information on the nature of work experiences that could then be matched to subsequent education choices and outcomes. The HS&B and NLS–72 data sets discussed above might be used for looking at the effects of work on lifelong learning education choices (e.g., determining who returns to what kind of schooling during their working life). Because the information on working during school is more limited, these surveys are less suited to secondary and more suited to traditional postsecondary education. Such information is best obtained from respondents who are still in school, ideally in secondary school. Existing surveys such as Beginning Postsecondary Students and Baccalaureate and Beyond are missing the secondary school experiences and, as such, are less than ideal.

The best approach is to start collecting data now on secondary school students—or perhaps even students in earlier grades—that will help us to understand how work affects education. Later on, the same data can be used to help understand how detailed educational experiences affect subsequent workplace success. The new data might include the following:

- Detailed information about work experiences during school of the kind noted above including the nature of the tasks performed, type of supervision offered, characteristics of training received, and so on. This information could then be related to subsequent academic achievement, course-taking patterns, and postsecondary experiences.

- More detailed information on KSAs including a student's work-based skills of the kind described by the SCANS report. The idea here would be to see how work experience affects these work-based skills. Later on, such information could be related to success in the workplace to see whether the results are different from those for academic achievement as more traditionally measured.

- Information on school-to-work programs and other work-based learning arrangements associated with schools. What effects do these arrangements have on academic achievement and on subsequent workplace success?

The School-to-Work Opportunities Office has funded two efforts to look at one aspect of these educational practices and arrangements. The first adds questions asking for details of school-to-work programs to the existing superintendent and school administrator's survey administered by the BLS. The second, more relevant here, adds questions to the National Longitudinal Survey of Youth about participation in such programs, information that can then be related to labor market outcomes. There may not be much of an argument for NCES to duplicate that effort with its own surveys. But when any of the existing NCES surveys are again in the field, adding even the same questions on participation in school-to-work programs would enable these surveys to examine the effects that participating in these programs might have on work outcomes. Similarly, the Bureau of Labor Statistics has proposed starting a new longitudinal study of 17-year-olds, and it is possible that this effort may also provide data to address some of the school and work questions.

Finally, there are many ways to collect data for research questions, and surveys of the kind at which NCES excels are obviously only one method for doing so. And it is probably worth a discussion as to what mix of survey data and other research approaches might be appropriate for addressing the questions described below. High-quality survey data with its enormous advantages in external validity are especially useful at capturing main effects of relationships between constructs that can be conceptualized and measured in a straightforward way. It is an important question as to whether selection issues and unmeasured attributes are intractable enough in some topics to demand more sophisticated experimental designs than are provided by national probability surveys. Whether surveys targeted toward particular populations might provide a middle ground between national probability surveys and experiments remains to be seen.

NOTES

1. It is worth pointing out that there is at least as much antipathy on the other side since many employers seem to distrust the goals that educators hold for students ("it's all about self-esteem, the kids aren't learning anything," and so on).

2. See Mishel and Bernstein (1994) for evidence on the former, and Gardner (1993) for evidence on the latter.

3. Much of the material in this section is drawn from Cappelli and Rogovsky (1995).

4. Preliminary findings from a study of transaction processing at the Wharton School's Financial Services Center find virtually no evidence of these practices.

5. These changes are described in Cappelli, P. (Ed.) *Change at Work*. (New York: Oxford University Press, forthcoming). A summary version of the arguments can be found in "Restructuring Employment," *Looking Ahead* (Washington, D.C.: National Planning Association, fall 1994).

REFERENCES

Adler, P. 1993. "Time-and-Motion Regained." *Harvard Business Review* 44(1): 97–108.

Altonji, J. 1995. "The Effects of High School Curriculum on Education and Labor Market Outcomes." *Journal of Human Resources* 30(3): 409–438.

Bartel, A., and Lichtenberg, F. February 1987. "The Comparative Advantage of Educated Workers in Implementing New Technology." *Review of Economics and Statistics* 69: 1–11.

Bowles, S., and Gintis, H. 1976. *Schooling in Capitalist America: Education Reform and the Contradictions of Economic Life.* New York: Basic Books.

Bray, D., Campbell, R., and Grant, D. 1974. *Formative Years in business: A Long-Term AT&T Study of Managerial Lives*. New York, NY: Wiley.

Bureau of the Census. "How to Enumerate the CPS." Washington, D.C.: 1989.

Campbell, J.P. 1994. "The Development of Occupational/Skill Clusters: Issues and Recommendations." Washington, D.C.: U.S. Department of Labor/Education and Training Administration.

Cappelli, P. April 1993. "Are Skill Requirements Rising? Evidence for Production and Clerical Jobs." *Industrial and Labor Relations Review* 46: 515–530.

Cappelli, P. 1995. "Changing Skill Requirements and Implications for the Structure of Wages" Paper presented at the Federal Reserve Bank of Boston Conference, November 17, 1995.

Cappelli, P., and Rogovsky, N. 1995. "Skills and Individual Performance." Working Paper. Philadelphia: National Center on the Educational Quality of the Workforce at the University of Pennsylvania.

Constantine, J.M. and Neumark, D. 1994. "Training and the Growth of Wage Inequality." Philadelphia: National Center on the Educational Quality of the Workforce (EQW).

Crawford, D., Johnson, A., and Summers, A. 1994. "School and Labor Market Outcomes." Working Paper. Philadelphia: National Center on the Educational Quality of the Workforce at the University of Pennsylvania.

Daniel, K., Black, and Smith, J. 1995. "College Quality and the Wages of Young Men." Working Paper. Philadelphia: The Wharton School, The University of Pennsylvania.

Demsey, R. September 1993. “An Occupational Classification System for Collecting Employment Data from Both Households and Employers.” In *Proceedings of the International Occupational Classification Conference*, pp. 235–250. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics.

EQW 1995. “First Findings: Results of the EQW/Census National Employer Survey.” Philadelphia: National Center on the Educational Quality of the Workforce EQW.

Gardner, J. June 1993. “Recession Swells Amount of Displaced Workers.” *Monthly Labor Review* 116(6): 14–23.

Hanushek, E. 1994, *Making Schools Work: Improving Performance and Controlling Costs.* Washington, D.C.: The Brookings Institution.

Ippolito, R. 1995. “Toward Explaining the Growth of Defined Contribution Plans.” *Industrial Relations* 34: 1–20.

Johnson, A., and Summers, A. 1993. “What Do We Know About How Schools Affect the Labor Market Performance of Their Students?” Philadelphia: National Center on the Educational Quality of the Workforce EQW.

Kane, T., and Rouse, C. June 1995. “Labor Market Returns to Two-and Four-Year College.” *American Economic Review* 85: 600–614.

Klein, J. 1989. “The Human Costs of Manufacturing Reform.” *Harvard Business Review* 42 (2): 60–66.

Marcotte, D. 1994. “Evidence of a Fall in the Wage Premium for Job Security.” Northern Illinois University, Center for Governmental Studies.

Medoff, J. 1993. “Middle-Aged and Out of Work: Growing Unemployment Due to Job Loss Among Middle-Aged Americans.” Washington, D.C.: National Study Center.

Mellow, W., and Snider, H. 1995. “Accuracy of Response in Labor Market Surveys: Evidence and Implications,” *Journal of Labor Economics* 1(4): 331–344.

Mishel, L., and Bernstein, J. 1994. *The State of Working America, 1994–1995.* Washington, D.C.: Economic Policy Institute.

Murnane, R, Willett, J.B., and Levy, F. 1995. “The Growing Importance of Cognitive Skills in Wage Determination.” *Review of Economics and Statistics* 77: 251.

National Commission on Education and the Economy. 1989. *America's Choice: High Skills or Low Wages!* Rochester, NY: National Commission on Education and the Economy.

National Commission for Excellence in Education. 1983. *A Nation At Risk: The Imperative for Education Reform*. Washington, D.C.: National Commission for Excellence in Education.

Osterman, P. 1995. “Skills, Training, and Work Organization in American Establishments.” *Industrial Relations*, April, 125–146.

Ree, M., Earles, J., and Teachout, M. August 1994. “Predicting Job Performance: Not Much More Than ˈgˈ” *Journal of Applied Psychology* 79: 518–524.

Rose, S. 1995. “The Decline of Employment Stability in the 1980s.” Washington, D.C.: National Commission on Employment Policy.

Sommers, D. 1993. “Ohio's Adoption of Canada's Jobscan Skill Checklists.” In *Proceedings of the International Occupational Classification Conference*, pp. 322–335. Washington, D.C.: U.S. Department of Labor, Bureau of Labor Statistics.

Stevens, D. W. April 1994. “Toward the Definition of New Occupational/Skill Clusters for the United States.” Washington, D.C.: U.S. Department of Labor.

Tyler, J., Murnane, R., and Levy, F. December 1995. “Are More College Graduates Really Taking ˈHigh School' Jobs?” *Monthly Labor Review* 118(12): 18–28.

Workforce 2000. 1985. Washington, D.C.: The Hudson Institute for the U.S. Department of Labor.

Discussant Comments

DAVID STERN

This is a lucid and lively paper. It builds on the substantial body of original research produced by the EQW Center, of which Cappelli is coordinator and to which he has contributed much impressive research of his own.

If I were to state all the points in the paper with which I agree, I would repeat most of the paper. Instead, I will select a few points to emphasize. And I will express a difference of opinion on one major issue.

Closer Connection Between Learning and Work

Cappelli is certainly correct that education *for* work, and especially education *through* work, have been relatively neglected topics in educational research. If we define education not as schooling, but as intentional learning, then the mere fact that the average person spends approximately 14 or 15 years in school but 40 to 50 years at work, engaging in some degree of intentional learning, should warrant greater attention to education after the end of formal schooling. This is all the more true because the degree to which work involves intentional learning appears to be increasing.

Cappelli describes how "high-performance" or "lean" production have broadened the responsibilities of front-line workers in manufacturing. Production workers have been called upon to learn quality control and job analysis. They are making changeovers to new products and learning new technologies at a faster rate, because their organizations must adapt or die. We are all caught up in accelerating change, born of faster computers, faster communications, faster flows of ideas and capital. This NCES meeting itself can serve as an example of education in the workplace as a response to these changes.

Cappelli also points out that "high-performance" management practices still do not prevail. Rather than investing in education for employees, many employers are choosing to "rent" people instead. The use of temporary staff has tripled since 1985, according to Cappelli. But temporary staff are continually learning, too: they are forced to do so as they move from one job to another. A study of Manpower Inc., the largest of the temporary staffing agencies, indicates that the company helps employees use their experience in a sequence of jobs to build a coherent portfolio of skills for themselves (Seavey and Kazis 1994).

The increasingly educative function of work is evident in the arrangements that some employers have adopted to promote learning. In addition to formal instruction in company

classrooms, many firms have devised methods of "just-in-time learning" that minimize the cost of learning by facilitating acquisition of skill and knowledge as part of the work process itself. Examples of these arrangements include cross-training of employees who work near one another, rotation of staff through a planned sequence of positions, and skill-based pay, which compensates individuals in part for what they demonstratably know and can do, independent of their specific job responsibilities during the pay period.

Researchers have debated whether or not changes in the workplace have resulted in a demand for higher levels of skill on the part of workers. Cappelli himself has produced some the most informative studies on this topic. However, as he explains in this paper, the definition of skill requirements is highly problematic. Procedures that personnel departments use to define skill requirements in practice are based on a priori judgments, not on demonstrated empirical relationships between skills and actual performance on the job. Cappelli would like to see more empirical validation studies of this kind. That is one of his main recommendations in his paper. However, I am less optimistic than Cappelli about the feasibility of mapping "KSAs" (knowledge, skills, and abilities) onto job performance. The plethora of distinct KSAs, and the multiplicity of job performance measures, make this research program at least as daunting as mapping the human genome—probably even more so, because job performance depends on contextual variables and its definition is constantly changing.

Instead of trying to specify KSAs and relate them to performance at work, it would be more feasible—and arguably more useful for policy—to test whether practices intended to promote learning at work lead to better performance by individuals or groups. I have mentioned some of these practices: cross-training, job rotation, and skill-based pay. These are all intended to promote the transmission of knowledge and skill from those who have it to those who want or need it. In addition, some organizations have procedures designed to promote the discovery of new knowledge in the work process. Cross-cutting this distinction, it may also be useful to classify workplace education practices by whether they take place "on-line" in the actual work setting or "off-line" in a classroom or other instructional milieu. This yields a four-way classification, examples of which are as follows:

	Transmission	**Discovery**
Off-line	classroom instruction	quality circles
On-line	job rotation, cross-training, skill-based pay	procedures to elicit suggestions for continuous improvement

Adult education surveys could include questions about participation in these and other arrangements for workplace education. In particular, teacher surveys could measure the prevalence

of these practices in their workplace, which is the school system. Further, the association between participation in such arrangements and the work performance of individuals or groups could be measured. If the study is longitudinal, it would also be possible to measure the correlation with performance in subsequent work settings for individuals who change jobs. Such studies would begin to illuminate whether and how education in the workplace affects performance at work.

How Work Affects Education

Cappelli also correctly emphasizes the fact that most students hold paid jobs while in high school or college (see also Stern and Nakata 1991). Indeed, the 1994 School-to-Work Opportunities Act encourages schools to incorporate more "work-based learning" into the curriculum. One logical justification for this policy is the expectation that students will become more capable of learning at work as adults if they practice doing it while in school. A study in France, where detailed statistics are collected on adult learning at work, indicates that individuals whose initial schooling included some work-based learning do, in fact, participate more in continuing education at work (Romani and Werquin 1995).

However, as Cappelli points out, most research in the United States on the effects of students' employment has considered only the amount of time they spend at work, ignoring qualitative aspects of their work experience. A recent exception is a longitudinal study conducted at the National Center for Research in Vocational Education, which has discovered correlations between certain characteristics of students' work and their school performance, as well as with their wages a few years later (Stone et al. 1991; Stern et al. 1995, and Stern 1996 forthcoming). NCES could build on this study to incorporate questions about students' job characteristics into longitudinal surveys of students, both K–12 and adult.

Conclusion

Traditionally, the connection between education and productive activity has been considered to be primarily sequential. Now it appears to be increasingly synchronous. NCES is in a position to provide essential data for describing and understanding the consequences of this convergence.

References

Romani, C. and Werquin, P. 1995. "Alternating Training and the School-to-Work Transition: Programs, Assessment, Prospects." Paper presented to the CERI/NCAL Roundtable on School-to-Work Transition in OECD Countries, Paris, February 2–3.

Seavey, D. and Kazis, R. 1994. *Skills Assessment, Job Placement, and Training: What Can Be Learned from the Temporary Help/Staffing Industry?* Boston: Jobs for the Future, Inc.

Stern, D., Finkelstein, N., Stone, J.R. III, Latting, J., and Dornsife, C. 1995. *School to Work: Research on Programs in the United States*. Washington and London: Taylor and Francis, Falmer Press.

Stern, D., Finkelstein, N., Urquiola, M., and Cagampang, H. March 1996 forthcoming. “What Difference Does it Make If School and Work Are Connected? Evidence on Cooperative Education in the United States.“ *Economics of Education Review*.

Stern, D. and Nakata, Y. 1991. “Paid Employment Among U.S. College Students: Trends, Effects, and Possible Causes.” *Journal of Higher Education* 62 (1): 25–43.

Stone, J. R. III, Stern, D., Hopkins, C., and McMillion, M. 1990. “Adolescents' Perception of Their Work: School Supervised and Non-School-Supervised.” *Journal of Vocational Education Research* 15 (2): 31–53.

THIS PAGE INTENTIONALLY LEFT BLANK

9 Using Administrative Records and New Developments in Technology

THIS PAGE INTENTIONALLY LEFT BLANK

Administrative Record Opportunities in Education Survey Research

Fritz Scheuren

INTRODUCTION

This paper addresses possible administrative record opportunities in the education survey research work of the National Center for Education Statistics (NCES). Elementary, secondary, and postsecondary education are included. The time horizon is roughly the next decade—through 2005—but some discussion will be provided that extends beyond that period, primarily in connection with the Decennial Census of 2010.

Organizationally, the paper is divided up into seven sections—(1) this introduction with some background and other introductory materials; (2) a look at overall trends that have an impact on administrative record electronic access; (3) possible scenarios in education statistics; (4) survey investment opportunities arising out of those scenarios; (5) related analysis opportunities and barriers; (6) the privacy and security issues that must be faced; and (7) a conclusion with an overall summary and some recommendations. An attempt has been made to keep the prospective broad, drawing on themes that are emerging or have emerged in statistical uses of administrative records generally. Naturally, there is particular emphasis on current NCES surveys.

Motivation and Goals

There is a widespread sense that the U.S. education system needs major improvement and that one way to help achieve this is through better statistical information systems (e.g., U.S. Department of Education 1991). NCES has produced an extensive array of survey products and related publications to address the need to monitor progress (see, especially, U.S. Department of Education 1994a). Many of the surveys it employs are based in whole or in part on administrative record data; however, still greater use of administrative records may be possible and it is the purpose of this paper to explore that option.

Scope of Administrative Records Examined

Formal administrative records in elementary and secondary schools and local education agencies are of six types:

1) Pupil records (cumulative folders, transcripts, etc.);

2) Instructional service records (courses offered, textbooks used, etc.);

3) Personnel records (specific teaching assignments, certification level, college transcripts, etc.);

4) Financial records (accounting journals, payroll records, etc.);

5) Records required by other agencies (health records, W-2s, etc.); and

6) Policy records (special tabular analyses and reports, etc.).

This list comes from a 1985 report prepared for the Office of Educational Research and Improvement (Hall et al. 1985). That source goes on to note that some records are initially maintained in separate school files and then summarized and entered into central record systems. The detailed content and organization of these files can vary from one local education agency to another—causing massive reporting and summarization problems as the records are further processed for use at the state level or for other entities, including NCES.

The information contained in these reports forms the core of a state's information system on local education agencies and schools. Supplementary data collections, including student testing programs, also occur. In 1985, though, for most states, these had not been integrated into a comprehensive educational information system.

Frankly, it is unclear as to the extent to which the above description of problems with administrative data continues to be true. There is evidence that matters have improved, with at least some states becoming highly automated (U.S. Department of Education 1994b). The overall extent of the progress being made is, however, unknown. Certainly, there have been some highly successful prototypes, notably in Nevada (Nevada Department of Education 1994).

The administrative records used for postsecondary education generically are similar to those for elementary and secondary institutions, with, of course, some important additions—like data from the federal student aid program. The impression is that colleges and universities, at least the large ones, are much further along in building integrated, electronic administrative databases.

One obvious recommendation to make for the future is to consider routinely and systematically tracking progress on improvements in the record management practices of at least a sample of the 15,800 local school districts and 85,000 public schools**.** Knowing how automation is proceeding in postsecondary institutions also should be routinely monitored, with success stories shared as appropriate.

In the next section there is a general discussion of overall trends that might affect strategies with respect to administrative record opportunities in NCES surveys. This is intended to frame the specific options that will be covered later.

SPECULATIONS ABOUT TRENDS

Making credible any prediction about the future is obviously problematic—especially involving technology up to a decade or more from now (Rennie 1995); hence, the approach here will be to discuss "scenarios," rather than to actually make any flat assertions about what will or will not happen. To motivate the scenarios to be discussed, broad, mostly obvious trends are speculated about below. These have been divided up into trends in computing, costs and budgets, survey science, institutional change, and concerns about personal privacy. What has been highlighted is not necessarily what is most likely to happen; indeed, in some cases, items are mentioned mainly because, if they did happen, great changes would have to be made in the way NCES currently does business.

Computer Technology

Among the possible computer technology trends (e.g., Ligon 1996) that bear on administrative record opportunities in education survey research are:

- Low cost personal and organizational computing power continues to spread ever more widely.
- Advances in telecommunications make possible the movement of increasingly large masses of data. The National Information Infrastructure effort is a major reason for this (Office of Science and Technology Policy 1994).
- Both of these trends are supported by increasingly powerful commercially available software.

The computing changes here are not only important in themselves, but they have opened up to many a whole new way to imagine the future. This has made people receptive to still other innovations too.

Costs and Budgets

A binding force that could harness these computer trends in a way that would increase information use of administrative records is what is happening to costs and budgets:

- Budgets could shrink greatly in all parts of the federal government (or at least not continue to rise).
- Costs of administrative data capture, in well-designed systems, would shrink too, perhaps at a rate faster than budget cuts.
- Costs of survey taking, on the other hand, have already dropped and would continue to do so, but at a slower rate (e.g., Nicholls 1988).

One implication of these observations is that administrative records will be much more available in an electronic medium than at present and at a lower and lower cost, relative to surveys.

This possibility is a central motivation behind any expansion of administrative record opportunities in education survey research.

Survey Science

The revolution that began just 100 years ago (e.g., Bellhouse 1988), with the advent of representative sampling, shows no signs of being overturned; nonetheless, the role of surveys and censuses could be modified greatly in the next decade (e.g., Scheuren 1995). Given what has already been said about structural changes in computing and costs, it seems possible that:

- Both novel and traditional uses made of administrative records will increase in lieu of surveys or in hybrid combinations. The 1996 National Postsecondary Student Aid Survey (NPSAS) is a clear example of what can be done in building a successful hybrid (see the Appendix).

- Surveys may play a "Rosetta Stone" role—to adjust administrative data and to help interpret such data, rather than being relied on directly to make estimates.

- Microsimulation and other modeling (e.g., National Academy of Science 1992) based on administrative data, statistically matched perhaps to outside sources, will increase—along with other prediction/projection techniques—because large-scale direct survey estimates might be affordable only at increasingly infrequent intervals.

Randomization-based survey estimates will continue to be the "Gold Standard" against which other methods of creating information will be calibrated. Budgets, though, will not permit the sample sizes of today and, as this paper argues, cheap administrative substitutes should be sought from which to make generalizations, especially for small domains and small areas (e.g., Boruch and Terhanian 1996; Schaible 1996).

Institutional Change

"Third-wave" ideas about how to organize and run educational and other large institutions may be widely tried (Toffler and Toffler 1994). This may span the gamut from an even more serious look at Japanese quality innovations (e.g., Mulrow and Scheuren 1995) to the breakup of public schools, as we now know them. If changes of this magnitude (e.g., Newmann and Wehlage 1995) get going during the coming decade, they can be expected to materially affect the incentives for providing access to administrative records. "Charter schools," say, could have the same costs and budget pressures as the elementary and secondary public institutions they replace. Universities may be the most affected, since their costs have risen the most steeply and since they may be altered the most (Noam 1995).

It is possible, however, that institutional changes could speed up rather than impede the other innovations envisioned above. Even so, "the breakup of the old order," should it occur, may cause real stresses in comparability of and access to administrative records across the nation's schools and colleges. Much the same comparability problem exists now though, and there are several public jurisdictions (parts of Maryland and New York, say) where cooperation already seems tenuous at

best—so what we may be dealing with here is more a matter of degree, rather than a major difference from the current situation (e.g., Salvucci et al. 1995).

Personal Privacy

Most of the trends mentioned above, arguably, could aid in increasing access by researchers to administrative records—at least not harming such an outcome greatly. Privacy and data security concerns, though, could slow down or permanently limit the growth in statistical uses of administrative records.

In some ways, NCES could not be better prepared to deal with such concerns. For example, the making and keeping of confidentiality pledges are nothing new at NCES; indeed, the Center has been an innovator in this area. Still the political debate coming could lead to legislation; predicting that law's effect on research will not be attempted here. Ways to mitigate any tradeoff between information needs and privacy exist though, and will need to be dealt with. For more on this, see U.S. Department of Education (1994b), where there is an extensive discussion of the early thinking of some of the state data stewards responsible for physical security and the protection of privacy. Use of the social security number, for example, is apparently already a sore point. It has been dropped from Virginia's student files. Moreover, access to any form of identifiable data, outside the local education authority, may be quite limited everywhere.

TWO SCENARIOS

Two scenarios are set out below. Each highlights what could be big changes from the current situation at NCES, relative to administrative records and their use with surveys. These scenarios are labeled, "Good" and "OK":

- "Good" is perhaps what one might want to happen.
- "OK" is a world that is livable but not desirable.

A "Poor" scenario was also looked at, but was so gloomy that it did not warrant writing about in detail.

The scenarios are both made up of a mix of the trends mentioned already, with a few natural extensions. As will be seen, there are common elements. Obviously, to the extent that the alternative futures set out here are credible, some of the commonalities noted may lead to anticipatory actions or investments on the part of the Center.

"Good" Scenario

Driven by concerns about international competitiveness (e.g., U.S. Department of Education 1994c) and the need to enhance the delivery of educational opportunities, a strong cooperative spirit

continues in the education community, even as the current institutional structures undergo change. Some elements of this "good" scenario are:

- Smaller survey budgets almost certainly will be in store for NCES, but cuts will be modest relative to cuts elsewhere in the federal government.

- Despite tight budgets, NCES' role as an information coordinator and catalyst will be desired and clearly recognized, in both the public and private sectors of education. An example of such support might be the statement by the Council of Chief State School Officers, namely "We strongly urge that . . . NCES be a true statistical center that assumes the major responsibility for coordination of the collection, assembly, analysis and dissemination for that sector of society under its purview, namely education." (Hall et al. 1985).

- Further, the changes envisioned should be gradual enough to allow NCES' electronic interchange efforts to link virtually the entire educational system into a common network. The National Research and Education Network (NREN) will establish a gigabyte communications infrastructure to enhance the ability of U.S. researchers and educators to perform collaborative research and education activities, regardless of their physical location or local computational and information resources. This infrastructure will be an extension of the Internet, and will serve as a catalyst for the development of the high speed communications and information systems needed for the National Information Infrastructure (Office of Science and Technology Policy 1994).

- This network could give ready access to identifiable electronically available administrative data at the school and maybe even at the student level (albeit this last is problematic, as noted earlier).

- Traditional, mainly paper school records would become increasingly automated, allowing for the education network envisioned to supply administrative data rapidly and cheaply for statistical uses.

- A flexible survey system, evolved from current NCES efforts, will make it possible to interpret these administrative data and to augment them when necessary.

- Samples are likely to be smaller in size than currently, but regular, with all the economies gained by continuous production and refinement. This may be a hard thing to sell, but a careful look at the time series versus cross-section tradeoffs (e.g., Ghosh et al. 1995) might make the case—especially if some of the administrative record proposals in the next section turn out to have value.

- Each "node" in the education network (state education office, school, or school district) will be able to create its own custom products; hence, the pressures of competition will work to keep the system innovative and cost effective. This third-wave approach brings each organizational element into the system, in some sense, as an equal.

- Standardization of administrative records will only be partial and full standardization may not even be seen as desirable by participants. That lack of full standardization obviously could be a major cost barrier, unless there is a change in underlying thinking about what the data mean. Again, this is a third-wave notion, but this time already well accepted in

accounting circles where each corporation—read school/district/state here (?)—can, within generally accepted accounting principles (GAAP), decide how to keep their books. Once established and approved, of course, the entity must continue in the same way. Why can't this work in education?

- Some elements of the educational system may be unwilling or unable to cooperate in sharing administrative data and so provision must be made for these. Groups that will require special treatment might be children being "home-schooled" (an already large movement that is likely to grow even larger). Some institutions of higher education (say, "Ivy League" schools) may also not want to be involved for other reasons. This is occasionally a problem already.

- Privacy issues will need to be carefully addressed; nonetheless, they should not be a major barrier to research uses of educational administrative records. Physical security and monitoring systems for administrative electronic data used in research will be a major cost of maintaining trust in the network being envisioned. Training to enhance "Privacy Literacy" among researchers will also be needed—again at no small cost.

- Use of administrative records from other systems should also be possible, including tax data (Forms 941, W-2s—maybe even 1040s), but access will necessarily be more limited—maybe only on a sample basis and with special consent arrangements. Partnering with the Bureau of Labor Statistics to use unemployment insurance records might also be possible; at least it should be explored.

Two last comments before going on to describe another scenario: (1) NCES may not be the only major information supplier in this networked world. Privatization is a distinct possibility. A lot will depend on how well the Center adapts to the changes coming. (2) It seems likely, though, that NCES could make "leading" contributions to developing the needed education information systems for this world. ("Leading" and "running" are not the same. It might be very undesirable for NCES to try to dominate in this networked world. To accomplish its mission, all it needs to be is a major player.)

"OK" Scenario

Again, as above, there is seen to be a compelling need by all to cooperate in achieving national education goals. More barriers to change exist, though, in this scenario, and a "limited success" is all that occurs in the coming decade. Some of the elements in this only "OK" world are:

- Declining survey budgets occur for NCES; the cuts, though, will be about the same as the average of cuts in statistical programs elsewhere in the federal government.

- Even so, a clear role for NCES as an information coordinator continues to be widely accepted, in both the public and private sectors of education. Resources to act as a catalyst in broadening administrative record research uses are, however, necessarily limited.

- Plausibly, the budget changes envisioned may not be gradual enough to allow NCES' electronic interchange efforts to link virtually the entire educational system into a common network. Still, most of the system could be networked anyway, but closer to the end rather than the beginning of the coming decade.

- This network would provide, as above, ready access to at least limited identifiable administrative data at the school but probably not at the student level.

- A flexible survey system, evolved from current NCES efforts, will make it possible to interpret these administrative data; for cost reasons, however, augmentation by direct surveying could be much less frequent than at present.

- Sample sizes are likely to be smaller as well, with few of the economies gained through continuous production and refinement.

- Some school or school district "nodes" will be able to create their own custom products, but this will not be an information-rich world—in many ways, information services may be about at the level they are today.

- Standardization of administrative records will be quite limited; however, developing and maintaining a metadata system, for at least the important concepts, should be attainable.

- Certain groups, like "home-schooled" children, despite their growing importance, will have to be ruled out of scope for most purposes.

- Privacy issues will need to be carefully addressed but still are not expected to be a major barrier to most research uses of available educational administrative records. Physical security and monitoring systems for administrative electronic data may be a concern in the network being envisioned, because only a "bare-bones approach" may be affordable. Training to enhance "Privacy Literacy" among researchers will have to be modest, exposing the system to a greater risk of a potential loss of trust on the part of the public.

- Use of administrative records from other systems could be very limited because of privacy and resource restrictions.

In summary, for this so-called "OK" scenario, NCES will at best be where it is today, except that inevitably budget cuts will have limited its information products at least somewhat. It is hard to imagine NCES leading, let alone running, the nation's education information systems in this world.

NCES Investment Opportunities

In the next section, we return to the overall trends mentioned earlier and suggest in broad terms what investments NCES might consider to increase the chance that the survey opportunities available in administrative records are enhanced—that is, that the "Good" scenario wins out over the only "OK" one.

SURVEY INVESTMENT OPPORTUNITIES

Administrative records play multiple roles in NCES surveys. Existing practice seems, therefore, to be a natural starting point for looking at further opportunities. Each of the major ways, current and proposed, where administrative records could be employed is discussed below, one at a time.

- *Administrative tabulations as a source of general information* are seemingly ubiquitous already. New opportunities here, if there are any, would lie in speeding up the availability of this information and potentially customizing it. On-line access is already fully in place for regularly prepared "ED TABS" summaries—e.g., as described in U.S. Department of Education (1994a); but see also what is being done elsewhere (Federal Committee on Statistical Methodology 1995).

- *Administrative data as a sampling frame* is very common too—at the school, teacher, or student level—e.g., in the public school components of the School and Staffing Survey (see McMillen, Kasprzyk, and Planchon 1993). Many opportunities exist, though, in this area. This is so especially if more data become electronically available on these frames, and quality improvements continue (e.g., Peng, Gruber, Smith, and Jabine 1993). Also, the time gap between the frame items and their potential survey use should be shortened; right now this can be up to 2 years or more.

- *Augmenting survey data with administrative items* during or after fieldwork is done in some NCES survey settings (e.g., NPSAS). Again, the opportunities for greater use of administrative records lie mainly in widening access to timely, electronically available data of high quality (U.S. Department of Education 1994c). Significant survey cost savings are obviously possible when comparable administrative data can be used, instead of obtaining the item by a direct survey method. The biennial NCES High School Transcript Study might be a place to begin to shift from the abstraction of data from paper records to direct electronic access. Differences in formats from state to state and even within states could be a major barrier, but a pilot might still be worth considering.

- *Editing survey data by comparing it to administrative items* is quite common in the establishment surveys of other agencies, such as Statistics Canada, the Bureau of Labor Statistics (BLS), or the Census Bureau. This use in NCES surveys seems to be infrequent at present, perhaps due to the timing and content of the administrative records that the Center has ready electronic access to. The Center has already sponsored studies (McMillen et al. 1993; Peng et al. 1993) which point to the possible benefits here, and pilot efforts to operationalize administrative data for editing survey variables might be among the steps to consider next.

- *Imputing for missing survey data using administrative records* is another common occurrence in establishment surveys at BLS and Census. Sometimes the administrative data are simply substituted directly; sometimes elaborate models are employed. It seems likely that both item and unit nonresponse (and perhaps coverage) adjustments could be improved if administrative data were employed. To test this idea out, NCES might want to conduct a pilot effort, say, with SASS and the Common Core of Data (CCD). This seems especially appropriate since so much analysis has been done recently with CCD and

SASS. Of particular note is that CCD is available every year. One year's CCD can be used, thus, as a frame while a later year can be used to edit the survey and impute for missing or erroneous entries.

- *Expanding the uses of administrative records that come from outside the education community* may be an important place to invest more. Privacy and security issues obviously are key here. Enhancing this option, through improvements in record linkage techniques, could even be a priority—especially for higher education, where IRS income data might become available because of the student loan program (National Academy of Science 1993). The ubiquitous social security number (SSN) seems the practical choice for student and teacher linkages, provided the SSN is backed up by confirmatory variables (such as names, addresses, and birth dates). School linkages to, say, Form 941 data or to unemployment records, should these be possible, would pose still other challenges.

  However, there may be a problem with this obvious approach, as already noted earlier, because of privacy considerations. Additionally, there are technical issues in the record linkage itself, especially without an exact identifier (e.g., Alvey and Kilss 1985; Newcombe 1988; Newcombe, Fair, and Lalonde 1992; Belin and Rubin 1993; Winkler 1995; Winkler and Scheuren 1995).

  Minor housekeeping improvements between NCES survey systems (and within such systems over time) might be looked at to see how broadly conformable they are to linkage, either using exact or statistical matching techniques (U.S. Federal Committee on Statistical Methodology 1980). The routine addition (or use) of check digits for all "unique number identifiers"—including for schools—is a suggestion for cases where they are not already on the survey or administrative records being employed by the Center. Achieving common formats for items that might be used to do statistical matching across administrative and survey systems also seems to be another option to look at.

- *Weighted survey estimates, obtained by poststratification to administrative totals*, might allow NCES to reduce current sample sizes and save money, without increasing the variance of major statistics. This could be done simply by employing conventional ratio estimation, using administrative data on the frame for both sampled and nonsampled cases. See Kaufman, Li, and Scheuren (1995) for more powerful and general methods too. Conceivably, even frame data that is a year or two old might be worth experimenting with. Better, more timely administrative data, of course, could lead to even better results.

- *Longitudinal surveys can particularly benefit from available related administrative data*. Administrative data can be used to help track cases (e.g., address changes) between interviews. Changes in administrative items may be predictive of similar changes in survey variables—among both respondents and nonrespondents. Clearly, editing and imputing longitudinal survey variables are greatly strengthened, if longitudinal administrative data have been linked. Times between successive interviews may be stretched out too, resulting in cost savings. Longer gaps between interviews, of course, would work only if the administrative data are near substitutes in the nonsurvey period. Staggered panels that have some direct data collection every year but at wider intervals

might be worth experimenting with, too, because of their potentially flexible, low cost nature.

- *"Mass imputation" of sample survey data to a complete population file* has been shown to work in some Canadian applications (e.g., Whitridge, Bureau, and Kovar 1990) and has advantages for NCES over simply weighting up administratively matched survey data. Mass imputation is a technique that assigns a survey case to one or more nonsampled cases in the population, using the overlapping data in some form of statistical matching. Each unit in the population is imputed a survey case. When efficiently done, the costs of mass imputing are only moderately larger than weighting. Recent work at the National Center of Health Statistics (NCHS) by Schafer and others in a Bayesian context provides an illustration of some of this method's real strengths—albeit for imputing for nonresponse (Schafer et al. 1993; Schafer 1991). Cheap computing is needed at the analysis stage because the whole population has to be processed. Given this last observation, it is not surprising that the Canadians, at only 1/10th the size of the U.S., were pioneers in this method. Nonetheless, the time is coming when the old computing cost barriers will be a thing of the past (even in government).

- *Mass imputation for small area estimates is also attractive* in an environment rich in detailed administrative data. Cross-section administrative data, like the Common Core of Data (CCD) for public schools, would be an ideal file to employ in experimental efforts to make small area estimates. To check this approach, a sample of areas—say, local school districts—would need to be selected. Direct survey observations in these selected areas would then be augmented sufficiently to test the idea. Obviously, for variables not closely related to those on the CCD not much should be expected—illustrating yet again the importance of expanding administrative items on NCES frames. The work NCES does with administrative records for small areas should, of course, not be confined to mass imputation, albeit mass imputation seems the most promising of the alternatives at this point (for more on small area estimation, see National Academy of Sciences 1992; Purcell and Kish 1979; Malec and Sedransk 1995; Schaible 1996).

- *Making survey time series estimates employing administrative data* is a natural extension of the methods being discussed. Initially, suppose that mass imputation techniques continue to be used. The step (leap) is from mass imputation (to cross-section administrative records) for small area estimates to doing mass imputation (to longitudinal administrative records) for time series estimates. Both start out with direct sample observations. In small area estimation a model is developed which predicts what the nonsampled cases would have reported in the survey for each element in the population in each area of the country. It is just one further, albeit big, step to predict what would have been reported by nonselected *and selected* cases, if the survey had been done again in, say, a different year. Obviously, changes in administrative data would be additional factors to consider in the imputation; that is, once an initial small area estimate had been made through imputation, it could be a starting point for small area *and time series* imputations for the next year. Time series estimation is an even older and deeper field statistically than is small area estimation; hence, other methods besides mass imputation

ought certainly to be tried. Whatever is finally done, the need to check on the estimates by direct survey measurement exists here too and could be a source of improvement ideas as well as helping to interpret the results.

- Administrative records, as indicated above, can be used by NCES in both novel and traditional ways. Some of these NCES has already been developing. In each example, though, the starting point was a survey. *What if the starting points were the administrative records themselves*, as is the case for most samples in some other agencies (e.g., the IRS)? In this later world, the main emphasis shifts to processing the administrative data and to using them directly for inference. Surveys could play a "Rosetta Stone" role—to adjust administrative data and to help interpret such data, rather than being relied on directly to make estimates.

At the outset of this paper, it was conjectured that randomization-based survey estimates would continue to be the "Gold Standard" against which other methods of creating information are calibrated. In this context, it was also said that future NCES budgets might not permit the sample sizes of today. Cheap partial (or complete) administrative record data might be appropriate substitutes, especially for small domains and small areas. As we have seen already, there are many ways for NCES to continue to take steps (big or little) in this direction. The implications of this "brave new world" for analysis, and analysts, will be covered next.

ANALYSIS OPPORTUNITIES AND BARRIERS

The previous section began with the existing ways that administrative records now support NCES surveys. Some ideas were also given on possibly strengthening these conventional methods. Gradually, though, the ideas for change moved more and more away from pure randomization-based survey inference; progressively, they were replaced by modeling ideas of various sorts (e.g., Särndal, Swensson, and Wretman 1991; Smith 1994).

Even supposing all of these ideas were sensible—and some of them undoubtedly will not work out—what would the benefits be? Is all this change worth the trouble? Is it possible that in order to save on data capture costs, other costs are being incurred that might be very large? Are costs being shifted from data producers to data users? Well, if a one-word answer were to be given, it would have to be "Yes"—at least some of the time. The old saw is also partly true, "We are trading the devil we know for the devil we don't." Unquestionably, one set of hard problems is being replaced by another.

Just look at the "Rosetta Stone" comment made above. While admittedly the most extreme of the options, this approach would be enormously challenging for educational researchers. In this world, surveys might be a much smaller part of the database, with many of the files being almost purely administrative. In such cases, survey vehicles would be used only to lightly monitor and interpret ongoing administrative data and to help explore new areas where administrative data did not yet exist—perhaps in an experimental setting or as part of an observational study of a new educational alternative.

This nearly completely administrative data world is not likely to happen soon—and for some information requirements, like opinion data, probably never. First, a much richer, fully networked, administrative data set is needed. Second, the eleven other options listed in the last section ought to be considered and maybe tried too—moving from those that are only modest extensions of what is now being done, to those requiring bigger and bigger changes on the part of both data producers and data users. Some additional steps are also recommended. Three of these are discussed in the subsections which follow.

Shifting the Emphasis From Data to Information

Understanding better the ways that current NCES data are turned into information by the Center itself, or through outside users, is an essential and obvious step. Data are products that, to be useful, must be "enlivened" by users. It is only through a positive synergy among data, data producers, and data users that information arises. Metadata systems are one of the best ways of making this synergy more systematic and more often fruitful. Strengthening Center efforts should be considered here, if only as a way of better tracking changes over which the Center has no control. NCES already does an outstanding job in running user training workshops and bringing interested individuals fully in contact with the data that the Center produces. Although already good, better file documentation is needed. Benchmarking studies on the metadata systems that other agencies (particularly administrative ones) are building might be a useful way to get potentially workable improvement ideas.

Further shifting of Center emphasis to providing information services rather than tabulations and data products cannot be stressed enough, as a way of preparing Center staff for the future discussed in this paper. Said another way: It is essential to look at the work being done from the customer end—realizing that all customers cannot be satisfied, even though that still should be the goal. Typically, data systems are very sluggish and change slowly. Information needs, on the other hand, move much more rapidly. A Center goal might be to develop *information systems* that are rapid, even though the *data systems* to which they are anchored may not be.

If, as seems likely, there will be more work for users to do as a result of the changes discussed in this paper, then one simple strategy is to *find more users to do it*. This admittedly "Tom Sawyer" approach is only a partial answer but it could help. Users have increasingly more powerful computing, possibly better than what NCES has, so big files and complex data structures may be seen as a welcome challenge to some—especially if there are more data overall and the data can be made more timely. In short, a marketing strategy might be warranted, and perhaps in market segments that are outside the traditional research community. With the proper privacy safeguards in place, these segments might include school administrators and other operating personnel (teachers and students?) at all levels of the national education system, who might want to compare themselves to those in similar circumstances. This expansion of users could go naturally, hand in hand, with a broadened access to secure administrative data for research purposes.

Getting the Distributions "Right"

Shifting to methods which emphasize more the need to get the "inference right," rather than just getting the "data right," seems essential. What does this mean? At present, most statistical agencies around the world spend a sizable fraction of their resources in collecting data and cleaning up inconsistencies in them—in short, on getting the "right data" (e.g., U.S. Federal Committee on Statistical Methodology 1990). Because these agencies are invariably peopled mainly by nonstatisticians, the idea that the data come from some underlying distribution with inherent uncertainties in it can get lost.

Technically, what is needed is to understand these distributions—to get them "roughly right," as Tukey has said. The data are thus only a means to an end, *not the end!* In a way, this is the same point made earlier, when information systems were being discussed. It is upon these distributions (Rubin 1990)—whether parametric or nonparametric, formal or informal—that inferences get made. The underlying causal mechanisms (or distributions) that generate the data observed are models that may, in the eye of the observer, be suggested by data or which can be fit to data. Distributions, thus, are a construct of the questioning observer. Obviously, the notion of distributions, then, unlike data, gives the user the central role.

"Selling" the user on data obtained from administrative records must be done for such records to be the basis for the creative leaps that research must make when new knowledge is borne. How might this be done? Assume two variables, one administrative and one survey-based, are compared and a scatterplot constructed that shows a strong relationship. Should the two variables be highly related, then arguably the same inference might be made from either one of them. Even so, the administrative variable might not be defined in quite the way that the researcher would like. On the other hand, the survey variable, while definitionally more suitable, could be costly to get; moreover, the survey variable would still be subject to sampling and measurement errors that could impair its use for inference. It truly is a question of deciding between the devil you know and the one you don't. Only experience will tell which devil is easier to live with. In any case, increasing reliance on administrative data may require experiments of the sort implied by this discussion. There is a lot at stake here. Put provocatively, should NCES invest in methods that may not even be based on exactly the "right data" but that could, most of the time, yield the "right inference" anyway? If the answer is "Yes," how might this be done? Beyond the answer, "it depends," not much of general value can be said here. Each such decision will need to be looked at individually.

Still, there is at least one comment worth making. With greatly expanded access to administrative data, the resources to do the careful (over)editing (Granquist and Kovar 1995) now characteristic of most survey systems would literally be impossible to find. Choosing new summary statistics that are robust against data problems is one obvious suggestion: medians instead of means, interquartile ranges in lieu of variances, graphical displays rather than tables of totals; all could allow users to see a distribution's shape in the presence of messy data. These or better methods make sense in the presence of administrative data of the scope envisioned. If the data suppliers are also data users (see "Shifting the Emphasis From Data to Information," above), then some of the Japanese quality improvement ideas might take stronger hold, leading to less back-end editing but without any sacrifice in "inference quality."

More Emphasis on Measuring Uncertainty

Strengthening the Center's efforts to measure sampling and other forms of uncertainty seems crucial too. At this point NCES has made great strides in building survey information systems that allow the user to measure sampling error. This is no small feat, given the complexity of the data collection. Much more will be needed, though, for the administrative data environment envisioned.

Some of the issues that will have to be addressed already exist today. For example, quantifying uncertainty in the presence of imputed data is an area of controversy at this point in surveys (e.g., Rubin 1996; Fay 1996). Mass imputation methods are not immune from criticism either (Rubin 1990). In a mass imputation world, of course, the administrative data would not be subject to sampling error. As far as the survey data go, they could have variances calculable, via methods that adjusted for the implicit poststratification that the imputation should generate (e.g., Wong and Ho 1991). How to estimate mean square errors for the joint distribution of survey and administrative data is an area that has been studied but seems to need more (basic?) research.

Among the tools being employed by NCES at present, resampling ideas, such as bootstrapping techniques, could be the best place to make further investments in estimating sampling variances (for more on bootstrapping in an NCES context, see, for example, Kaufman 1995). Gibbs sampling tools could help, too, if more general measures of uncertainty were desired.

Winners and Losers

In this section, three analysis issues have been briefly discussed in the context of a possible large-scale expansion of administrative record use—with or occasionally in lieu of NCES surveys. The topics covered were illustrative and not exhaustive:

- To focus more on the information end, rather than the data end of the Center's work;
- To reallocate resources away from data cleaning[1] and toward better ways to see underlying distributions; and, finally, and very briefly
- To look hard at techniques to measure uncertainty that work during the period when the transitions envisioned will be taking place.

Clearly, if and when the most radical of these administrative record changes came about, there would be major consequences for education researchers. Since the time span is so long—10 years or more—and given that small experimental intermediate steps are possible, adjustment problems seem manageable. This is not to say that adjustment will be easy; in some places they can be predicted to be hard indeed.

PRIVACY, CONFIDENTIALITY, AND DATA SECURITY

Privacy, confidentiality, and data security issues have been given considerable attention in many forums in recent years. The range of treatments is a wide one, spanning the 1993 book, *Private Lives and Public Policies*, which focused on research data and was intended for specialists, to the very recent book, *The Right to Privacy*, which, while also a considerable scholarly accomplishment, is intended for a more general audience (Duncan, Jabine, and de Wolf 1993; Alderman and Kennedy 1995).

Tore Dalenius has provided a good review of privacy, confidentiality, and security goals in statistical settings. His work may afford a point of departure here (e.g., Dalenius 1988; see also Boruch and Cecil 1979). In common speech, the words privacy, confidentiality, and security partially overlap in usage and often have meanings that depend greatly on context. Each can also have an emotional content which makes precise definitions difficult, even contentious. For example, Dalenius quotes Westin (1967) about privacy: "Few values so fundamental to society as privacy have been left so undefined in social theory or have been the subject of such vague and confused writing by social scientists."

A good start on giving meaning to the word "privacy," or "information privacy" (our context here), might be the definition first articulated by Justice Brandeis as the "right to be left alone . . . the most comprehensive of rights and the right most valued by civilized man" (Olmstead 1928). With books like *Private Lives and Public Policies*, it appears we may finally be making serious progress in operationalizing the Brandeis definition of "privacy rights"—at least as they relate to statistical information. Much remains to be done though. The practices of data stewards in education (U.S. Department of Education 1994b), and elsewhere (Jabine 1993) vary widely. Public opinion research shows a range of concerns, too, depending on the context in which questions about privacy are asked (Scheuren 1985, 1995; Blair 1995; Presser and Singer 1995). Information on informed consent exists too but is dated (Singer 1993).

The National Center for Education Statistics has been the pathbreaker in giving controlled access to its survey files for qualified researchers (e.g., Wright and Ahmed 1990). The Center needs to continue taking the same kind of leadership position with regard to assuring wide educational research access to administrative records as it has with surveys (e.g., U.S. Department of Education 1995).

The final outcome here, though, is quite uncertain, since each state may legislate separately on the kind of electronic access that will be permitted for statistical purposes. Identifiable school level data are already extracted (in CCD). Having identifiable student level administrative data at NCES would be desirable, for example, for many surveys too. Overall data security issues deserve NCES attention, particularly as electronic administrative data become more and more widely available. In this regard, the recommendations[2] in the report *Educational Data Confidentiality* (U.S. Department of Education 1994b) are worth quoting at length:

> There appears to be a need to inform those who work with electronic data and citizens as well as taxpayers of laws, regulations, and procedures that schools, states, and regional agencies adhere to in collecting, using, and protecting data confidentiality. Such information should be widely available, readable, and easily

understood. It should summarize current federal and state assurances of privacy and limits on data access and use, and be accessible to the public through government agencies at local, state, and federal levels. These central findings are suggested:

- Standards, procedures, and recommendations are available from other agencies, and from states that have established workable procedures, but there is relatively limited cross-agency or cross-state exchange, and wider dissemination of models would advance the security of new systems.

- States and other data agencies should be encouraged to inform agency personnel who work with personal record information—including student records, personnel records, and family demographic information—what regulatory restrictions limit access and use and encourage staff persons to make an effort to keep members of the public well informed of these rules, assurances, and routine protections of privacy.

- States, districts, and other data agencies need more routine procedures for publicizing widely across agencies and among taxpayers and citizens the confidentiality protections they have in place.

Some areas where emerging issues may need monitoring are mentioned below. It should be noted, that while this list has many challenges, it is by no means exhaustive. These are:

- How to manage the physical data security for this new information network, so that the system is fully "auditable"—i.e., access records are kept of what was looked at, by whom, what changes are permitted and get made (for more here, see Brannigan and Beier 1995).

- How to assure that proper notification and consent procedures are followed so that individual human rights are respected (e.g., Singer, Shapiro, and Jacobs 1995; Scheuren 1985; Scheuren 1995). Continuing experiments seem the wisest course here and might be worthy of consideration by NCES.

- How to adjust for cases where consent is denied to administrative records by NCES survey respondents. This is a very tough problem if the refusals are at all sizable, which does not seem likely at this point. Basically what seems needed is to institute statistical work on group matching or other techniques that would lessen the tradeoff between the competing values of furthering scientific research *and* safeguarding personal privacy (e.g., Spruill and Gastwirth 1982; Gastwirth and Johnson 1994).

- How to track public opinion on the education research uses being made of the linked data network being built. The series of Harris-Equifax surveys are one source here, albeit imperfect (Harris et al. 1993). The Harris-Equifax surveys have important limitations (Blair 1995) on their interpretability; nonetheless, their main conclusions are in essential agreement with other research on privacy concerns. Roughly, almost no matter how you ask the question, there are always about one sixth to one fifth of the population who oppose electronic record linkages on privacy grounds. Conversely, again almost no matter

how you ask the question, about the same fraction will favor "beneficial sounding" linkages on efficiency grounds. The two thirds or so in the middle will differ in their opinions depending on the specifics. See also Presser and Singer (1995).

- How to protect research data from nonresearch uses, especially by governmental entities. See, especially, Chapter 1, *Private Lives and Public Policies* (Duncan, Jabine, and de Wolf 1993).

- How to reduce inadvertent reidentification risks, especially those that arise through school level linkages with student data. This is the same problem, in some ways, that exists with the Social Security Administration's Continuous Work History Sample (CWHS). See, for example, Jabine and Scheuren (1985).

Clearly, privacy and related confidentiality and security issues must continue to be faced as administrative data become increasingly available electronically. The uncertainties about the future seem greater here than elsewhere but the Center has done a lot already and seems poised to do more.

SUMMARY AND RECOMMENDATIONS

In this paper, there has been a broad discussion of opportunities for making more effective use of administrative records in surveys of elementary, secondary, and postsecondary education. Here it may be appropriate to group what has been said concerning:

- *How is the availability of data in administrative records likely to change during the next decade and how will these changes influence opportunities for NCES data collection and analysis?* The first three sections of this paper cover topics in this area. Predicting the future is so difficult that two scenarios were used and formed the basis of discussion. Many specific suggestions were made in passing. The underlying premise, though, as far as basic research is concerned, is that the Center probably cannot afford to make major financial investments. Staff investments are needed, nonetheless, in monitoring the changes coming and mining them for ideas to try in ongoing Center efforts. The Center's role as a technology transfer catalyst is where investments should be made, if possible, in bringing the good ideas on-line faster. It might be necessary to help bring cheaper administrative data capture and electronic transfer technologies to schools so that they can lower or at least contain these "back-office" costs—much as banks and insurance companies have begun to do. What, for example, can the Center do to help create and test cheap scannable forms for some routine transactions and bankcard-like direct electronic access for others?

- *What are the opportunities for better integrating surveys of individuals (for example, students, teachers, administrators, or parents) with existing administrative records to improve the quality and utility of NCES surveys?* Here the Center can and needs to do the most. Imbedded experiments with new methods of design, data capture, estimation, and analysis should be a growing part of NCES survey efforts. A whole range of these was discussed under "Survey Investment Opportunities" and "Analysis Opportunities and Barriers," above. In particular, work involving CCD is a natural place to make a

concentrated effort (Holt and Scanlon 1994), but following up on the design ideas in the 1996 NPSAS makes a lot of sense too (see Appendix). Center tradition and recent research supports such growth. There is, however, the usual problem with all surveys of being conservative about change, once a survey has begun to operate (Dillman 1994; Groves 1995). Looking at survey contract vehicles and staff incentives will be crucial to overcome the natural risk adverse behavior that is likely to exist.

- *What are the main issues or barriers surrounding access to or better use of administrative records, and how might these be addressed?* Throughout the paper, but especially in the "Privacy, Confidentiality, and Data Security" section above, there were places where the many barriers to change were dealt with. Some of these can be overcome by gaining new knowledge, e.g., by more methods research on, say, CCD—perhaps an experiment to directly access school administrative records, rather than continue to transcribe them as at present.[3] For many issues, a wait and see approach may be the only strategy possible. Especially for changes in institutional arrangements, the Center probably has no role, except to react to events. There are still activities to be considered, however. For example, developing generalized capabilities to react is one option here. In the context of administrative record access, for example, in the privacy area continuing to work toward a fully secure network environment for research, auditable by each school and even each student, could go a long way to overcome potential concerns.

Still another activity that might be emphasized, in the Center's applied research, is the private school segment of elementary and secondary education. An extra effort in this area would warrant consideration, depending on what seemed the likely speed of movements to change the "Old Order," such as the creation of Charter schools. What about experimenting with partial Internet-available (encrypted?) administrative record alternatives to the Private School Survey?

AN AFTERWORD

Of course, even if the Center does not try to speed up beneficial change, change is inevitable and, hence, the Center will have to deal with imbedded experiments involving all sorts of changes, including to administrative records. The question is what role will NCES take in their design or even whether they get designed or just happen. Crucial, too, is how will the Center protect its surveys when these experiments go wrong, as occasionally might occur, no matter how well they are designed.

One of the most encouraging things is that those who welcome the future changes coming will not be alone. Virtually all large organizations are moving in the same direction (Nanopoulos 1995), even statistical ones (e.g., Keller 1995). The positive synergy from the massiveness of what is happening should sweep up those organizations, like NCES, who want to change and move them much farther than their individual efforts alone would make possible.

The way these outside changes play out within the educational community may deserve the most staff attention. As noted already, especially important from an administrative record perspective is speeding up the automation and networking of administrative records, since without broadened access and drastic cost cuts, such records will largely remain on paper and, hence, hard to obtain for survey research.

One final point, however optimistic one may be about the (distant?) future, there is a long way to go. The *Wall Street Journal*, in a special section on school (mainly computer) technology, dated November 13, 1995, made this point extremely well (*Wall Street Journal* 1995; see also *Science* 1995). The Center has, though, clearly made a good start. It is hoped this paper will help too.

NOTES

1. This observation may seem in conflict with one of the recommended additional uses of administrative records (i.e., for editing) discussed in the previous section. The issue is not to stop editing, but to stop overediting, a point made in the 1995 Granquist-Kovar paper cited earlier.

2. Also of interest are the views expressed in the report of the Privacy Working Group on the U.S. National Information Infrastructure (1995).

3. At present, the CCD is not processed as a longitudinal file. Longitudinal (transaction-based) processing is another important improvement to consider, especially anticipating the day when administrative data are put on the CCD directly without a separate extraction step.

REFERENCES

Adelman, C. 1995. *The New College Course Map and Transcript Files: Changes in Course-Taking and Achievement, 1972–1993* (based on the Postsecondary Records from Two National Longitudinal Studies). Washington, D.C.: U.S. Department of Education, Office of Educational Research and Improvement, National Institute on Postsecondary Education, Libraries, and Lifelong Learning.

Alderman, E., and Kennedy, C. 1995. *The Right to Privacy*. New York: Knopf.

Alvey, W., and Kilss, B., eds. 1985. *Record Linkage Techniques—1985.* Washington, D.C.: Department of the Treasury, Internal Revenue Service.

Belin, T., and Rubin, D. 1993. "A Method for Calibrating False-Match Rates in Record Linkage." *Journal of the American Statistical Association* 90: 694–707.

Bellhouse, D. 1988. "A Brief History of Sampling Methods." In *Handbook of Statistics.* Elsevier Science. Representative sampling seems to have first been proposed in an 1895 paper by Kiaer. This was a major advance, as routine as such surveys seem today. Kish, in writing a paper on the history of surveys, captures the real battle (which continues) in the title of his presentation, "The Hundred Years War of Survey Sampling," which he delivered in Rome on May 31, 1995.

Blair, J. 1995. "Ancillary Uses of Government Administrative Data on Individuals: Public Perceptions and Attitudes." Unpublished Working Paper. Washington, D.C.: Committee on National Statistics, National Academy of Sciences.

Boruch, B., and Terhanian, G. 1996. "Trends in Statistical and Analytic Methodology: Implications for National Surveys." Paper presented at the Conference on Future NCES Data Collection: Some Possible Directions, Washington, D.C., November 1995, and included in this volume.

Boruch, R., and Cecil, J. 1979. *Assuring the Confidentiality of Social Research Data.* Philadelphia: University of Pennsylvania Press.

Brannigan, and Beier. 1995. *Medical Data Protection and Privacy in the United States: Theory and Reality.* Paper submitted with support from American Association of the Advancement of Science, Science and Human Rights Program.

Cecil, J. 1993. "Confidentiality Legislation and the United States Federal Statistical System," *Journal of Official Statistics* 9 (2): 519–536.

Czaijka, J., Moreno, L., and Schirm, A. 1995. *On the Feasibility of Using Internal Revenue Service Records to Count the U.S. Population.* Report prepared for the U.S. Department of Treasury, Internal Revenue Service by Mathematica Policy Research, Inc.

Dalenius, T. 1988. *Controlling Invasion of Privacy in Surveys*. Continuing Education Series, Statistics Sweden.

Dillman, D. 1994. *Why Innovation is Difficult in Government Surveys.* Presented at the 1994 Census Bureau Annual Research Conference.

Duncan, G., Jabine, T., and de Wolf, V. 1993. *Private Lives and Public Policies.* Washington, D.C.: National Academy Press. See also the review by Scheuren, F. March 1995. *Journal of the American Statistical Association*.

Fay, F. 1996. "Alternative Paradigms for the Analysis of Imputed Survey Data," *Journal of the American Statistical Association.*

Gastwirth, J., and Johnson, W. 1994. "Screening With Cost-Effective Quality Control: Potential Applications to HIV and Drug Testing," *Journal of the American Statistical Association* 89: 972–981.

Ghosh, D., Smith, W., Chang, M., and Saba, M. 1995. *Optimizing the Periodicity of the Schools and Staffing Survey: An Interim Assessment Based on 1987–88 and 1990–91 Data.* Washington, D.C.: U. S. Department of Education, National Center for Education Statistics.

Granquist, L., and Kovar, J. 1995. "Editing of Survey Data: How Much is Enough?" Paper delivered at the International Conference on Survey Measurement and Process Quality, Bristol, UK, April 1995. Incidentally, one way to partially avoid the problem of overediting is provided in Lawrence, D., and McDavitt, C. 1994. "Significance Editing in the Australian Survey of Average Weekly Earning," *Journal of Official Statistics* 10: 437–447.

Groves, R. 1995. "Challenges of Methodological Innovation in Government Statistical Agencies." *The Future of Statistics, An International Perspective.* Ed. Z. Kenessey. International Statistical Institute.

Hall, G. et al. 1985. *Alternatives for a National Data System on Elementary and Secondary Education*. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Harris et al. 1993. "Harris-Equifax Privacy Survey" (for 1993 and earlier). New York: Louis Harris and Associates.

Holt, A., and Scanlon, B. 1994. *QED Estimates of the 1990–91 Schools and Staffing Survey: Deriving and Comparing QED School Estimates with CCD Estimates.* Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Jabine, T. 1993. "Statistical Disclosure Limitation Practices of United States Statistical Agencies," *Journal of Official Statistics* 9 (2): 427–454.

Jabine, T., and Scheuren, F. 1985. “Goals for Statistical Uses of Administrative Records: The Next Ten Years.” *Journal of Business and Economic Statistics*.

Kaufman, S., Li, B., and Scheuren, F. 1995. “Improved GLS Estimation in NCES Surveys.” Paper delivered at the meetings of the American Statistical Association, Orlando. Incidentally, this paper forms a partial pilot for the poststratification ideas discussed here.

Kaufman, S. 1995. *Properties of the Schools and Staffing Survey's Bootstrap Estimator*. Paper presented at the meetings of the American Statistical Association, Orlando.

Keller, W. 1995. “Changes in Statistical Technology.” *Journal of Official Statistics* 11 (1): 115–127.

Keller-McNulty, S., and Unger, E. 1993. “Database Systems: Inferential Security.” *Journal of Official Statistics* 9 (2): 475–500.

Ligon, G. 1996. *New Developments in Technology: Implications for Collecting, Storing, Retrieving, and Disseminating National Data for Education*. Paper presented at the Conference on Future NCES Data Collection: Some Possible Directions, Washington, D.C., November 1995, and included in this volume.

Malec, D., and Sedransk, J. 1995. “Small Area Inference for Binary Variables in the National Health Interview Survey.” Washington, D.C.: U.S. Department of Education, National Center for Health Statistics.

McMillen, M., Kasprzyk, D., and Planchon, P. 1993. “Sampling Frames at the United States National Center for Education Statistics.” *Proceedings of the International Conference on Establishment Surveys*, 237–243.

Mulrow, J., and Scheuren, F. 1995. *Measuring to Improve Quality and Productivity in a Processing Environment*. Paper presented at the Orlando meetings of the American Statistical Association. This is an application of Japanese, or Deming-like ideas, to surveys—focusing on cultural and organizational change.

Nanopoulos, P. 1995. “Expected Changes in Record Keeping,” In *The Future of Statistics, An International Perspective*. Ed. Z. Kenessey. International Statistical Institute.

National Academy of Science. Committee on Applied and Theoretical Statistics. 1992. *Combining Information: Statistical Issues and Opportunities for Research*. Republished in *Contemporary Statistics* #1, American Statistical Association, 1992. This is a good general reference to the broad area of combining information. Many of the hybrid ideas covered in the section on Survey Investment Opportunities in this paper fall under this general rubric. Virtually every federal agency involved in assessing the impact of alternative policy proposals uses models depending on micro-data: information relevant to highly aggregate (e.g., national) conclusions, but measured at a much lower level of aggregation.

National Academy of Science. 1993. Student Aid Program.

Nevada Department of Education. 1994. *S.M.A.R.T. Plan, Statewide Management of Automated Record Transfer: A Plan to Automate and Transfer Student Records Statewide*. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Newcombe, H., Fair, M., and Lalonde, P. 1992. "The Use of Names for Linking Personal Records," *Journal of the American Statistical Association* 87: 1193–1207.

Newcombe, Howard B. 1988. *Handbook of Record Linkage*. New York: Oxford University Press.

Newmann, F., and Wehlage, G. 1995. *Successful School Restructuring.* A Report to the Public and Educators by the Center on Organization and Restructuring of Schools, University of Wisconsin, School of Education.

Nicholls, W. 1988. "Computer-Assisted Telephone Interviewing: A General Introduction," In *Telephone Survey Methodology*. Eds. R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg. New York: Wiley, pp. 377–402. See also Batcher, M. and Scheuren, F. 1996. CATI site management in a survey of service quality, *Bristol Survey Quality Measurement Conference.* Wiley: New York. In the literature on computer- assisted surveys, usually the observation is made that data capture costs do not drop over what were, heretofore, the conventional methods. Because of the higher quality obtained, however, it could be argued that total costs have fallen.

Noam, E. October 1995. "Electronics and the Dim Future of the University," *Science* (270): 247–249. The university has been based on a "model—centrally stored information, scholars coming to the information, and a wide range of information subjects housed under one institutional roof—logical when information was scarce, reproduction of documents expensive and restricted, and specialization low. It became also the model for the most formidable of knowledge institutions of antiquity, the Great Library of Alexandria." These conditions are no longer true and, as a consequence, major changes seem in store.

Office of Science and Technology Policy. 1994. *High Performance Computing and Communications: Toward a National Information Infrastructure.* A Report by the Committee on Physical, Mathematical and Engineering Sciences.

Olmstead v. United States. 1928. 277 U.S. 438–478. (Justice Brandeis dissenting).

Peng, S., Gruber, K., Smith, W., and Jabine, T. 1993. "Monitoring Data Quality in Education Surveys." *Proceedings of the International Conference on Establishment Surveys*, pp. 244–252. Some of these are also discussed in the NCES *Working Paper Series*. Washington, D.C.

Presser, S., and Singer, E. 1995. "Public Beliefs about Census Confidentiality." Preliminary Report based on part of the 1995 Joint Program in Survey Methodology Practicum Survey.

Purcell, N., and Kish, L. 1979. "Estimation for Small Domains," *Biometrics* 35: 365–384.

Rennie, J. September, 1995. "The Uncertainties of Technological Innovation." *Scientific American, Key Technologies for the 21st Century, 150th Anniversary Issue*. According to Rennie, "Few of the promised technologies failed for lack of interest." One recurring reason that predictions were off "is that even the most knowledgeable forecasters are sometimes much too optimistic about the short-run prospects for success." He then adds, "The more fundamental problem with most technology predictions . . . is that they are simplistic and, hence, unrealistic. Sadly, some inventions are immensely appealing in concept but just not very good in practice."

Rubin, D. 1996. "Multiple Imputation After 18+ Years." *Journal of the American Statistical Association.*

Rubin, D. Discussant remarks in the *1990 Proceedings of the Annual Research Conference, (pp. 676–679).* Washington, D.C.: Bureau of the Census.

Salvucci, S., Zhang, F., Monaco, D., Gruber, K., and Scheuren, F. 1995. "Multivariate Modelling of Unit Nonresponse for the 1990–1991 Schools and Staffing Surveys." Paper delivered at the American Statistical Association Meetings, Orlando.

Särndal, C.E., Swensson, B., and Wretman, C. 1991. *Model Assisted Survey Sampling*. New York: Springer-Verlag. For a criticism of such methods for a statistical agency, see Smith, T. 1994. "Sample Surveys: An Age of Reconciliation, 1975–1990." *International Statistical Review* 62: 5–34.

Schafer, J. 1991. "Algorithms for Multiple Imputation and Posterior Simulation from Incomplete Multivariate Data with Ignorable Nonresponse." Unpublished Doctoral Dissertation, Department of Statistics, Harvard University.

Schafer, J., Khare, M., and Ezzati-Rice, T. 1993. "Multiple Imputation of Missing Data in NHANES-3." *Proceedings of the Annual Research Conference* (pp. 459–487). Washington, D.C.: Bureau of the Census.

Schaible, W. ed. 1996. *Indirect Estimators in U.S. Federal Programs.* New York: Springer-Verlag.

Scheuren, F. 1985. "Methodological Issues in Linkage of Multiple Databases." *Record Linkage Techniques-1985*. Washington, D.C.: Department of the Treasury, Internal Revenue Service.

Scheuren, F. May 1995. "An Administrative Record Census in the U.S.?" *Chance*. The costs of primary data collection for statistical purposes are simply not competitive with an administrative source, if such a source can be made to serve. Why capture the same, or almost the same, data more than once? This point has been made over the last 15 years with regard to census taking. The 2010 Decennial Census will almost certainly be based in large part on administrative records, as is already common for many European countries. Education records could be among those used; much will depend on how the 2000 Census research program is carried out.

Scheuren, F. March 1995. "Review of Private Lives and Public Policy." *Journal of the American Statistical Association.*

Scheuren, F. 1995. *Linking Health Records: Human Rights Concerns.* Paper submitted with support from American Association of the Advancement of Science, Science and Human Rights Program.

Scheuren, F., and Li, B. 1995. *Intersurvey Consistency in NCES Private School Surveys.* Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Science 1995. "Getting Wired." *Random Samples* 270: 1929.

Singer, E., Shapiro, R., and Jacobs, L. 1995. "Privacy of Health Care Data: What Does the Public Know? How Much Do They Care?" Paper submitted with support from the American Association for the Advancement of Science, Science and Human Rights Program.

Singer, E. 1993. "Informed Consent and Survey Response: A Summary of the Empirical Literature." *Journal of Official Statistics* 9 (2): 361–375.

Spruill, N., and Gastwirth, J. 1982. "On the Estimation of the Correlation Coefficient from Grouped Data." *Journal of the American Statistical Association* 77: 614–620.

Toffler, A., and Toffler, H. 1994. *Creating a New Civilization*. Atlanta: Turner Publishing. The authors provide a set of questions that allow one to determine whether an approach is "third wave." Among other things, a third wave approach should 1) not resemble a factory—i.e., not be built on such principles as standardization, concentration, and bureaucratization; 2) not massify society—i.e., focus on the individual; 3) not have too many eggs in one basket.—i.e., push as many decisions as possible down from the top and out to the periphery; and 4) not be vertical but be virtual. Contrast this with the centralized vertically integrated survey system proposed for NCES in Hall, G. et al. 1985.

U.S. Department of Education. 1993. *SpeedeExpress, An Electronic System for Exchanging Student Records*. Washington, D.C.: Council of Chief State School Officers, National Center for Education Statistics.

U.S. Department of Education. 1995. *Advanced Telecommunications in U.S. Public Schools, K–12*. Washington, D.C.: National Center for Education Statistics.

U.S. Department of Education. 1994b. *Education Data Confidentiality: Two Studies, Issues in Education Data Confidentiality and Access and Compilation of Statutes, Laws, and Regulations Related to the Confidentiality of Education Data*. Washington, D.C.: National Center for Education Statistics, National Forum on Education Statistics.

U.S. Department of Education. 1994a. *The Condition of Education 1994*. Washington, D.C.: National Center for Education Statistics. This publication may be a direct response to the need to monitor progress toward the goals mentioned in the report *Education Counts: An Indicator System to Monitor the Nation's Educational Health.* 1991. Moreover, the *Condition of Education* report series is, of course, only one of many general and specialized NCES publications that illuminate this important national priority.

U.S. Department of Education. 1994c. *Methodological Issues in Comparative Educational Studies: The Case of the IEA Reading Literacy Study*. Washington, D.C.: National Center for Education Statistics. For a somewhat different perspective, see Rotberg, I. December 1995. "Myths about Test Score Comparisons." *Science* 1: 1446–1448.

U.S. Department of Education. 1991. *Education Counts: An Indicator System to Monitor the Nation's Educational Health*. Report of the Special Study Panel on Education Indicators to the Acting Commissioner of Education Statistics. Washington, D.C. Among the goals to be monitored for the year 2000 are 1) All children in America will start school ready to learn; 2) The high school graduation rate will increase to at least 90 percent; 3) American students will leave grades 4, 8, and 12 having demonstrated competency in challenging subject matter including English, mathematics, science, history, and geography (and leave school) prepared for responsible citizenship, further learning, and productive employment in the modern world; 4) U.S. students will be first in the world in science and mathematics achievement; 5) Every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship; and 6) Every school in America will be free of drugs and violence and will offer a disciplined environment conducive to learning.

U.S. Federal Committee on Statistical Methodology. 1995. "Electronic Dissemination of Statistical Data." *Statistical Policy Working Paper 24.* Washington, D.C.: Statistical Policy Office, Office of Information and Regulatory Affairs, Office of Management and Budget.

U.S. Federal Committee on Statistical Methodology. 1990. "Data Editing in Federal Statistical Agencies." *Statistical Working Paper No. 18.* Washington, D.C.: Office of Management and Budget.

U.S. Federal Committee on Statistical Methodology. 1980. *Report on Exact and Statistical Matching Techniques*. Washington, D.C.: U.S. Department of Commerce.

U.S. Information Infrastructure Task Force. 1995. *Privacy and the National Information Infrastructure: Principles for Providing and Using Personal Information.* A Report of the Privacy Working Group. Washington, D.C.

Westin, A. 1967. *Privacy and Freedom*. New York: Atheneum.

Whitridge, P., Bureau, M., and Kovar, J. 1990. *Mass Imputation at Statistics Canada.* U.S. Bureau of the Census, Sixth Annual Research Conference.

Winkler, W. 1995. "Matching and Record Linkage," in *Business Survey Methods*. Eds. Cox, Binder, Chinnappa, Christianson, Colledge, and Kott. New York: Wiley.

Winkler, W., and Scheuren, F. November 1995. "Linking Data to Create Information." Paper delivered at the 12th Annual Statistics Canada Methodology Conference, Ottawa.

Wright, D., and Ahmed, S. 1990. "Implementing NCES's New Confidentiality Protections." *1990 Proceedings of the American Statistical Association,* Section on Survey Research Methods. Alexandria, Va.

Wong, W., and Ho, C. 1991. "Bootstrapping Post-Stratification and Regression Estimates from a Highly Skewed Distribution." *1991 Proceedings of the American Statistical Association,* Section on Survey Research Methods. Alexandria, VA.

APPENDIX

National Postsecondary Student Aid Study by Dennis Carroll

In 1996, the fourth National Postsecondary Student Aid Study (NPSAS) will collect information from all types of students in all types of postsecondary institutions. This study, which is about two-thirds the size of previous administrations of NPSAS (1987, 1990, and 1993), builds upon the collection strategies of the earlier studies and incorporates administrative records from the major student financial aid programs. NPSAS is a comprehensive study, spanning aided and unaided students, independents and dependents, employed and non-labor force participants, undergraduates and students seeking advanced degrees, as well as full-time and part-time students. There are three major users of NPSAS data: NCES, USED, and financial aid policy analysts. NCES uses NPSAS to profile groups of students (e.g., undergraduates, part-timers, minorities, borrowers), and NPSAS serves as the base year for longitudinal studies (i.e., BPS and B&B). USED (including PES and OPE) used NPSAS to determine the rates of receipt of federal student financial aid for various subgroups of students. Policy analysts describe aid issues and build models.

NPSAS is an extremely complicated set of six integrated data collections: enrollment lists, Central Processing System (CPS) records, Computer Assisted Data Entry (CADE) for institution records, Computer Assisted Telephone Interviews (CATI) for students, CATIs for parents, and Pell grant/loan award files. From a sample of all (IPEDS) institutions, lists of all students (undergraduate and graduate/first-professional) enrolled at any time during the academic year (July 1–June 30) are collected (with business majors flagged, if possible). For the longitudinal component, lists of first-time students (BPS) and filed for graduation students (B&B) are collected and unduplicated. The initial NPSAS sample of students is selected from these lists.

The USED CPS contains application and preliminary award information for federal student aid, in particular, the Pell grant program. The initial NPSAS sample is matched with the CPS information to obtain data on family finances and preliminary awards of federal aid. (CPS does not contain all federal aid. Much of the federal loan program data are currently fragmented in several files, which may be consolidated in the next several years.) Matched CPS data are preloaded into the CADE.

To extract data from institutional records housed in student financial aid offices, admission offices, and graduate dean's offices, a CADE is used. Many institutions (over 60 percent) complete the CADE data collection on their own, but some institutions do not have staff or computers. Contractor staff travel to these institutions and complete the CADE for them. The CADE extends CPS data to gather the core NPSAS information on student aid—including all federal aid, state aid, institutional aid, and assistantships. In addition, information on program (e.g., intensity, major, admission, and demographics) is collected. The CADE data are preloaded into CATI systems.

The CADE data allow subsampling of students who did not receive student financial aid. (About 40–45 percent of students receive aid.) In NPSAS:96, the initial sample of 59,000 has been reduced to a more efficient sample of 37,000 by undersampling unaided students.

The two CATI systems for students and some parents require location and collection systems. Only a subgroup of parents are interviewed consisting of mostly dependent and unaided students' parents. If in tracing the student the parent is contacted, then the parent CATI is conducted; otherwise, the parent CATI follows the student CATI. The parent CATI gathers data on family finances that parallel the CPS information.

The student CATI expands the NPSAS data to cover six areas: other aid, non-school costs, labor force activities, family structures and finances, future plans and goals, and community activities. Other aid covers small programs that do not flow through student financial aid offices and aid from other institutions attended during the academic year. Non-school costs include living expenses, transportation, and child care. Labor force activities include employment (sometimes in college work-study program jobs), program related employment (internships), and lack of employment. Family structures and finances include marriage, children, and other dependents; the earnings/assets of the household; and the expenses of the household. Future plans and goals include occupational, community, and personal aspirations. Community activities include citizenship and service.

Finally, the Pell grant and federal loan award files (in their final audited form) are merged.

NPSAS yields three recurring policy reports, two data analysis systems, and a restricted set of data files for secondary analyses. The recurring policy reports are *Profile of U.S. Undergraduates*, *Financing Undergraduate Education*, and *Financing Graduate/First-Professional Education*. Separate data analysis systems are built for undergraduates (about 700 variables) and graduate/first-professionals (about 600 variables). Finally, the data files (with the associated methodology report) are made available to licensed users for secondary analyses (including Postsecondary Education Descriptive Analysis Reports [PEDAR]).

New Developments in Technology: Implications for Collecting, Storing, Retrieving, and Disseminating National Data for Education

Glynn D. Ligon

OVERVIEW

The National Center for Education Statistics (NCES) is seeking a future vision for data collection, storage, retrieval, and reporting. This vision will guide improvements in data collection and reporting processes to increase the availability and usefulness of data while decreasing the burden on local and state agencies. This paper describes the developments in technology that will affect the collection and reporting of education data. A major implication is that information solutions present challenges that are as much human resource issues as technology issues. The lack of acceleration in our use of technology is attributable in large part to the shortage of individuals trained and capable of making the technology work, within an environment that encourages the use of technology. For NCES, staffing roles, responsibilities, and skills must change along with the introduction of technology solutions.

Summary of Implications for NCES

NCES should position itself to ride the wave of automation in the nation. The trends described herein are as follows:

1) Faster computers will allow NCES to expand the amount of data collected, analyzed, and reported while potentially reducing the time and burden imposed on reporting agencies and NCES staff.

2) Increased storage capacity on computers will allow NCES to collect and maintain as much data as is reasonable to collect based upon the information needs of audiences.

3) The universality of networks will allow NCES to collect data electronically, communicate to clients electronically, and make available its analyses and reports electronically.

4) EDI standards and software will make electronic data exchanges over these networks efficient, effective, and affordable.

5) Relational data base concepts will be applied to a distributed information system that will allow access to data across agencies' files.

6) Productivity software will automate information management tasks to the extent that staff will insist upon computer applications over any remaining manual processes.

7) NCES can achieve the benefits of an individual student-level database without the problems of creating one within NCES. Emerging networks and data standards can create a national distributed information system. NCES would be able to query each state database to conduct analyses without having to maintain individual records centrally.

The NCES Data Warehouse

Technology supports NCES's plans to develop a data warehouse. A data warehouse is simply a location where someone can access information electronically. The NCES data warehouse should be a library containing both books with statistics and analyses already accomplished and raw data available for analysis.

Criteria for Judging the Future System

"Alternatives for a National Data System on Elementary and Secondary Education," 1985, proposed a set of criteria to be used for judging a national education information system. These criteria are applied to the vision described here. In a reverse of position from 1985, confidentiality will move from the bottom to the top of the list of concerns requiring careful attention by NCES.

Conclusion

Ensuring that NCES's data collection, storage, analysis, and reporting processes take full advantage of technology will be a process, not an event. This transition will require considerable training and support for both NCES staff and the staff of its data providers. When evaluated against the criteria described in 1985, the vision of the future as described here would be a significant improvement over past and current systems.

INTRODUCTION

The National Center for Education Statistics is seeking to establish a vision for data collection, storage, retrieval, and reporting for the future. This vision will guide the planning and implementation of improvements in the data collection and reporting processes. This effort is significant for many other agencies beyond NCES. Nationwide, decision makers, parents, educators, students, businesses, and others are affected by the availability and quality of education data. The expectation is that technology advances will provide opportunities for solutions that will increase the availability and usefulness of data while achieving decreases in the burden imposed upon local and state agencies to collect and report the data.

Technology is already available to support the processes described. In fact, the NCES staff have already used some of the newer methodologies on a limited scale. A challenge will be to escape

the inertia of traditional systems, to create a new inertia of change, one that shortens the time required to go to scale with technology-enhanced solutions.

This paper describes the developments in technology that have affected or will affect the collection and reporting of education data. The underlying premise is that we must have a vision for a new national education information system. Our vision, based upon function, needs to drive our decisions and actions. At present, many decisions are reactions to new technology as it is developed. We are intrigued by technology and want to adapt our needs to it. An important perspective for us in the education information arena is that our needs should inspire a search for technology that provides solutions to those needs. The functional aspects of our data collection and reporting should change as our needs change. Technology is one direction in which to look for solutions to our changing needs.

A major implication from the discussion in this paper is that the information solutions to be explored and implemented present challenges that are as much human resource issues as technology issues. For education institutions, the reality has been that the capability of technology is ahead of the capability of individuals to apply that technology to our information systems. In other words, much of what educators are asking to do now can be accomplished with existing technology or straightforward adaptations of hardware and software. The lack of acceleration in our use of technology is attributable in large part to the shortage of human resources, individuals trained and capable of making the technology work, within an environment that encourages the use of technology. For NCES, staffing roles, responsibilities, and skills must change along with the introduction of technology solutions.

Importantly, the future system described in this paper must be responsive to issues raised in the other papers in this series. The technology used in our future information systems must be chosen because it is responsive to the demands detailed in the other papers. A major concern within education agencies today is the purchasing of hardware and software because they are available and appear to be useful. The alternative is to seek hardware and software solutions for problems that have been clearly identified. This is particularly evident in the instructional arena where the users may be inspired by technology, but not have the time and resources to integrate it adequately into their processes.

For NCES and the future of education information at the national level, a major hurdle with which to contend is the variety in both type and age of the technology that must be integrated across schools, districts, postsecondary institutions, states, and NCES to create a functional information resource for decision makers, parents, businesses, educators, staff, and others. Schools, districts, and state agencies have been acquiring technology (e.g., computers, printers, modems, and so on) since the early 1980s. Much of that hardware is still in use, irrespective of how out-of-date it has become. Some states purchased hardware when large sums of dollars became available to their legislatures. Since that time, dollars to upgrade have been more difficult to find. So, on the one hand, this paper makes the point that hardware is relatively inexpensive to purchase now. However, on the other hand, available funds for purchases may be scarce.

As examples, consider the situations in South Carolina and Georgia. South Carolina raised taxes for education once about a decade ago and purchased that era's state-of-the-art hardware and

a student information management software system for schools. Now the software is being updated by the vendor, requiring a newer, more powerful operating system that will not run on the old hardware. About 7 years ago, the Georgia Legislature approved funds to build a student information system across all schools. Over those years, hardware purchases have been made to bring about 70 percent of the schools to an operational status. However, those schools where implementation occurred years ago have old hardware compared to the schools being brought on board this year.

Another perspective is found in Texas. Almost a decade ago, plans to build a statewide information database were begun. At the time, accommodations were made in the design for schools that were still punching 80-column cards. Currently, the 80-column format is still being used for the computer files submitted. Other states are designing information systems now that are incorporating relational database designs to be much more efficient. The dollars and human resources required for Texas to reengineer its existing system are huge.

Technology may be capable, but are the users in education agencies ready? Is the technology present in the education arena? How out-of-sync will agencies become as the financially advantaged acquire capabilities that others do not have—or as agencies replace and upgrade at varying rates? Do we have to plan for the lowest common denominator?

We need to build a vision of functions, not of hardware. We need to envision systems whether or not the infrastructure exists to support them, then we need to build toward that vision, ensuring that each step taken is consistent with the long-range goal. Space travel has taught us that all the pieces of technology do not have to be in place before a project can begin. New techniques and products can be developed along the way.

Data Collection

Data quality must be achieved and maintained throughout all areas of information systems, but it begins with adequate standards during collection. There is a balancing between timeliness and quality that threatens to undermine the ultimate purpose of data collection, which is to inform decision making. Data quality must become a priority for the future. Information systems must be designed to provide timely education data that can be used with confidence as the basis for decisions.

The mechanics of data collection are changing already—from paper-and-pencil forms to optical scanners, to computer screen entry, to disk exchanges, to electronic file transfers, to direct reading of distributed files, to simultaneous updating of remote files as transactions occur. All of these are existent to some degree across educational agencies. Our vision must motivate agencies to continue moving up the hierarchy of automation. As data are collected at the local and state levels in automated fashion, they are more readily available for exchanging up the system to NCES. The vision for data collection must include the idea that redundant, independent data collections will be coordinated. Changes in retrieval and access processes as described below allow for collection of data from files within databases rather than requiring that someone reformat the data to fit a forms-based report.

Data Storage

Data storage media are increasing in capacity and decreasing in cost. The changing formula of cost-per-piece of information stored indicates that we can and will allow ourselves to be less disciplined about what we store and how long we maintain it. The implication is that more unrefined, raw data will be maintained and be available for analyses. With faster processors, more individuals will have the ability to process huge files of raw data. Already we have seen the discipline of sampling theory decline in importance in research. More studies are conducted on population statistics rather than sample statistics, because the cumbersome calculations required for sophisticated statistical procedures are handled easily by computers. Advances have changed how we store images, translate voice to text, scan text and translate it to word processing files, and create documents and data files without ever producing a paper document. This paper discusses the implications for coping with a data system that grows to include so many elements in so many formats. The emerging methodology for data warehouses will provide some answers.

Data Retrieving

Retrieving, which will also be thought of as access, is the function that supports the utility of data and makes it more valuable. In the automated world, the separate concepts of retrieving and disseminating begin to blur. As audiences gain access to data, the act of someone disseminating the data is no longer necessary. Retrieving and disseminating can be viewed as all being part of a single process that makes data available to users in a wide range of states of development from raw to fully analyzed. This paper discusses how future information systems will employ a range of access techniques to accomplish retrieval/dissemination. Access will be closely linked to issues of security, confidentiality, and integrity of both the data themselves and the analyses and conclusions drawn from them. This will be a controversial issue. Determining who can access which data elements within a database will be difficult. Controlling access to ensure that only those authorized to access certain data are allowed to will be an even bigger challenge.

"Regulated access" allows the owner of data to place them in a location for access, without requiring that owner to package and send them to every requester. Today, someone within an organization typically prepares a response to an information request and sends it. With regulated access, the owner of the data will monitor who is accessing and using them rather than providing the data directly. Requesters/readers have responsibility for establishing their credentials for access and usage.

A CHANGE IN PARADIGMS

Although overused today, the phrase "changing our paradigms" applies precisely to the automation of data collections and the use of the resultant data. An important concept in this change will be that the nature of data collection will evolve. We will not want to merely automate manual or paper systems. When conversions are made to technology-based systems, the process underlying the collections should change to take full advantage of how the technology operates.

Some other ways our thinking must change are as follows:

1) Survey forms *will be replaced by* data files that do not look at all like the paper surveys.

2) Dissemination of reports *will be replaced by* interested audiences accessing information in electronic form and printing the parts they want.

3) Statistics calculated and published by a single agency *will be replaced by* competing statistics calculated from the same database by both private and public entities.

4) Keeping all the data you need on your own computer *will be replaced by* networked databases from which your computer can access huge data sources.

5) A computer programmer responding to a request for a report *will be replaced by* having the person who needs the report run it.

6) New mandates requiring new data collections *will be replaced by* new mandates resulting in an analysis of data from an existing, shared database.

7) Data burden being defined as the amount of time required to document activities and complete reports *will be replaced by* its being defined as the overwhelming amount of data available for consideration.

8) Statistics and reports being published months after collection *will be replaced by* immediate access to data as soon as they are uploaded to a central file.

Within the context of its charge to provide useful and timely statistics about education, NCES is finding that many other agencies and organizations collect and report data as well. Professional organizations survey members and the general public often these days. Commercial polling services conduct numerous, seemingly continuous, surveys of public opinion. With the expansion of computer storage capacity and the move toward providing public access to data and report files, there arises the issue of how much of these related data collections should be acquired and made available by NCES.

Several issues are clear. First, does NCES endorse or make an implicit statement about the quality of other organizations' data by redistributing them? What obligation does NCES inherit when it redistributes these data? Secondly, is this redistribution necessary? As will be described in this paper, the technology allows for NCES to point audiences to other information sources using electronic connections without having to copy the data they are seeking onto an NCES computer. The cautious approach would be to leave data collected by other organizations and agencies on their own information systems and resolve the technology issues of how to connect potential audiences to them as appropriate.

WHAT ARE THE DEVELOPMENTS IN TECHNOLOGY THAT AFFECT EDUCATION DATA SYSTEMS?

Advances in technology are very technical and complex within the covers of our computers and other hardware. However, to the users of information systems, the relevant aspect of these advances is function. Function can be described as the operational actions that a user notices. What does the application do for you? How well does it do it? How fast does it perform that function? What manual activities are replaced? When microprocessing chips are miniaturized through amazing advances in manufacturing, the end user notices that computers grow smaller and faster at the same time. When modems advance in their transmission speeds, the end user notices that activities that used to take too long to be practical over a phone line can now be accomplished reasonably. So, the technical advances that result in faster chips and modems are discussed here more in terms of the impact they have on users. The impact on users translates directly to implications for the next generation of NCES data collection and reporting systems.

Developments and their implications for NCES are discussed within these areas:

Hardware: The physical items that make up the computer and its visible components

Network: A group of two or more computer systems linked together; the telecommunications systems that link computers

Software and Applications: The instructions that tell the computer what to do

Hardware

Storage Capacity

Compared to the 1980s, today's data storage devices present fewer limitations on the quantity of data we can have readily available to us. A storage device is the object onto or into which data are placed. These include hard disk drives (internal or external magnetic disks); removable floppy diskettes, cassettes, or cartridges using magnetic disks or tape; and optical disks (compact disks or CDs). For comparison purposes, commonly found hard drives of under 50 megabytes in the 1980s would not even hold some of today's data files that can exceed 100 megabytes for elaborate publications with graphical images. As this paper is being written, families are buying 1 gigabyte hard drives for their homes. The floppy disks of a decade ago have been replaced by removable disks and cartridges that hold several gigabytes of data.

Storage capacity is not limited by the advertised level on floppy disks, tapes, and cartridges. The demand for affordable, large storage has inspired software developers to design data compression routines that remove all the unnecessary bits of information out of a file. These compression routines can achieve impressive results, such as reducing the space required to store a file by 10 to 90 percent.

The impact of these advances in storage efficiency is that the limits are being removed on the number of files and the amount of data that can be maintained within an individual's computer. When NCES began keeping the many statistics it collects from the states on computer files, the size of those files was a major consideration, and the cost to add more storage to hold more files and data had real budget impact. Today, several hundred dollars can solve a large data storage need. The direction of technological advances continues to be toward greater and greater storage capacities, in less and less space, for fewer and fewer dollars.

For NCES, this means that constraints that used to be placed upon expansion of data files and conversion of paper records to an automated format have faded. NCES is capable of holding within local computers virtually all the data that are practical to collect and enter. Future decisions determining the data to be collected can be made upon need and usage factors rather than available storage capacity.

In the past, researchers were required to understand and use sampling theory to create reliable data sets for analysis. With limitations upon the ability to access data on mainframe systems or to store large data sets on personal computers, a premium was placed upon collecting manageable sample data sets. Considerable professional literature has been produced to guide researchers in this process. Probability statistics have been common in the literature to provide readers with an understanding of how much confidence they should place in the findings of studies. Educational research is now using population statistics from large databases that include measures of every individual of interest. The constraint is more on the collection methodology (how practical it is to measure every individual) than on the data storage and analysis capacities. This trend will be evident in the future operations of NCES. As a data warehouse is built and stocked, more and more data will find its way into it. Fewer and fewer restrictions will be imposed based upon lack of storage space.

Another aspect of data storage that has changed involves the benefits from expanded electronic networks. With a local area network installed, NCES can store data on multiple computers throughout the agency and create an environment that functions as a virtual single source for data. This concept also works on a much broader scale outside NCES. Any agency that shares a common set of standards for exchanging data files can be a part of a distributed information system. Such a system would allow sharing of data while maintaining internal integrity and local control. This would be in contrast to a true distributed database within which all agencies must comply with exactly the same data definitions and formats. Those implications and benefits are discussed in more detail throughout this paper. The bottom line is that in a networked environment, the users have virtually on their own desktops the data from all computers linked by the same network.

Telecommunications Speed

In the 1950s, “faster than a speeding bullet” (miles per hour) impressed us because it was too fast for us to actually see. In the 1990s, data traveling virtually at the speed of light carries our communications over fiber optics. The result is that we no longer describe the efficiency of a computer or the transmission of data as “how fast something is moving.” Our data transmissions have reached a plateau in how fast they move. Speed is now defined in terms of how much

information can be sent from one place to another within a certain amount of time (bits per second). The bullet Superman outraced traveled intact, arriving at its destination in the same physical shape as it left. Data files are stretched out and arrive literally in bits and pieces. In telecommunications, the goal is to send and receive as many data in as little time as possible. In other words, the performance goal is to stretch out the data file as little as possible so the first bit that arrives is followed as soon as possible by the last one. This goal has been pursued with great success. Most casual personal computer users have noticed that their modems (the devices that translate information into and out of the characteristics required for transmission) evolved very quickly from 1200 baud (roughly 1,200 bits per second) to 2400, 9600, 14.4 (notice the change in notation to units of 1,000 with 14,400 being expressed as 14.4), to a common modem on store shelves transmitting at 28.8. A 28.8 baud modem sends about 24 times as much data as the old 1200 baud modem did in the same amount of time. This miracle is achieved in great part through eliminating any unnecessary bits of information in a data file and compressing everything into as few bits as possible to carry the same meaning when decompressed at the other end.

What implications does speed have for the future of NCES data collection and reporting? Faster telecommunications will allow for larger data sets to be exchanged efficiently. Again, this removes a barrier to designing future systems. Future information systems will not have to be constrained as much by the time and expense factors in data exchange. NCES can collect more data in large data sets without imposing a greater burden on states and others in terms of transmission time and costs. Today, NCES's trading partners are already finding it to be more practical to extract and transmit data electronically compared to copying data onto a floppy disk and physically sending it. Across the state education agencies, few have not implemented some data submissions on disk, and some have implemented submissions over networks.

Processing Speed

Another speed issue is how long a computer takes to perform the millions of transactions it is asked to do for a specific application. Processing speed is one of the more difficult concepts to discuss. There are many factors that determine actual processing time for a computer task, e.g., access time for storage devices, input/output time for other components of the computer system, and the amount of time required for the monitor to recreate images as they change. Even the casual personal computer user knows that the speed at which personal computers' central processing chips perform tasks has increased dramatically. Miniaturization in the manufacturing of chips continues to progress. Simply put, tasks that took hours in the 1980s were reduced to minutes in the 1990s, and are now being completed in seconds.

The implication for NCES is similar to that for all education researchers. We can now calculate complex analyses on large data sets within a more reasonable amount of time. As discussed earlier, the need for sampling strategies and sampling statistics is reduced. A researcher can use an entire data set on a population of individuals. For future planning, NCES does not have to be so concerned with having large data sets to analyze and the burden that places on staff and the time that requires to publish statistics. The option presented to NCES will be to produce more and more analyses and reports within the same amount of time, or to publish the same analyses more quickly—or both.

Access Speed

When speed is discussed, there is another component beyond the central processor and the modem that has an impact on how quickly an individual can accomplish work on a computer. A major factor is access speed for all the storage devices. The access speed determines how long it takes the computer to move data from the storage device into its active memory (random access memory or RAM). Data must be in RAM to be processed. Larger computer programs require that data be moved into and out of RAM periodically. CD-ROM players moved from single-speed access to 4x, or four times, the speed for the original cost within about 2 years.

Improvements in access speed contribute to the overall performance of computer systems. Again, the limitations on future information systems of NCES are shrinking. The task of maintaining and using a very large information system is taking less time.

Random Access Memory (RAM)

RAM is the random access memory a computer uses to keep data readily accessible for processing. A useful analogy is the human brain. The brain stores tremendous amounts of memories. We could never keep all those memories active in our conscious at one time, so only that information that is needed for thinking at any one time is called upon. The computer calls up those data it needs for the current task it is performing into RAM. The greater the capacity of a computer's RAM, the more information that can be kept handy for processing at one time. Commonly installed RAM has grown from 1 to 2 megabytes 3 years ago to 4 to 16 today. Newer operating systems (the essential directions that tell the computer how to run software programs) require greater RAM. This trend appears to be a given to continue or even to accelerate.

Another counterbalancing trend is the increased usage of RAM and storage capacity by newer operating systems. The implication of this is that as operating systems (e.g., Windows 95) improve, they will require more RAM to operate and more disk space to be loaded. The future of prices for RAM is uncertain, so it is not possible here to predict whether the increased RAM required in the future will cost more than the amount required in today's machines.

There is also a benefit for large information systems. Computers with adequate RAM will perform large, complex tasks quicker. This is one area where added productivity comes at a cost. The installed computers in many offices are old enough to have inadequate RAM to run the newest operating systems and applications.

Printers and Graphics

A brief note is appropriate here about the visual appeal and communicability of the output from the newest publishing/printing systems. A desktop computer can now produce the impressive color graphics that once were the sole venue of professional layout artists and printers. For NCES, the benefit is that staff can make publications more reader-friendly and more likely to be read.

Network

Up front, we should recognize that easier access to networks has been a priority feature of newer operating systems (e.g., OS/2 Warp, Windows 95). The user's challenge to learn how and to take advantage of networks becomes easier with each new generation of operating systems.

Local Area Networks

Computers within a single location can be connected to each other to share resources in a local area network (LAN). Physically, a LAN consists of a card inside each computer, wires between the computers, and network software to manage the communications between the computers. Printers and other devices may also be connected through the LAN. Some LANs are this simple. Others can use wireless communications, multiple access units and routers to direct the transmissions between locations, and servers. Servers are computers that store data and software, and manage the operations of the LAN.

LANs expanded in the late 1980s as users discovered the advantages of sharing printers, using electronic mail, working on the same documents, and reading data on another computer. A single user gained the power of several computers. With the recent installation of a LAN within NCES, this potential is available to staff.

Wide Area Networks

The Internet is a wide area network (WAN). WANs connect computers that are located in separate places. LANs may be connected by WANs. The distinction between a LAN and a WAN is the amount of separation between the computers. However, the technical requirements, legal parameters, and operational issues for a WAN are much more complex than for a LAN that is self-contained within a single location.

The Internet is a public network that connects anyone to anyone else who chooses to connect. Public institutions, including state education agencies and postsecondary institutions, are almost universally connected. Across these agencies, the level of usage varies. However, NCES currently has access to its major data trading partners through the Internet. School districts and schools are connecting quickly. However, some are far from being automated in their operations, and some of those choose not to be for the foreseeable future. Therefore, NCES can assume that the Internet is available for use by its primary information trading partners, but that those partners may be exchanging information with others who are not connected to the Internet.

Although not free as is commonly thought, the Internet is relatively inexpensive to connect to and use. The Internet is far from simple to access within some agencies. In 1995, NCES and the Office of Migrant Education sponsored a pilot across six sites to use the SPEEDE/ExPRESS standards as a basis for the exchange of education records for Migrant Education Program students. The expectation was also that the solution for migrant students would apply as well to all mobile students, who make up about 20 percent of students annually. Each volunteer site was to be connected to and using the Internet as a prerequisite for participation. The reality was that one site

had only personal accounts used by a few staff members, another had no connection, another was connected, but required a multistep process for the Migrant Program staff to be trained and issued an address, and another used a gateway to a university Internet provider that required changes in the EDI software being used to connect. The other two sites were in Florida, which has an established statewide network. However, the Internet connection was set up through their state-level office rather from each district. The pilot demonstrated that the logistics of actually using the Internet for data exchanges can be much more involved that some may think.

Value Added Networks

Value added networks (VANs) are the private enterprise equivalent of the Internet. Although structured very differently, the functionality of VANs and the Internet are similar. Customers pay a VAN for usage of their network services. The value-added aspect is that the VAN provides services and features that the Internet expects the individual users to take care of themselves. The features include controlling access to users, guaranteeing connections, and providing some degree of security.

Very recently, VANs began making connections to the Internet available to their clients on a limited basis. Although too early to count on, the trend is for VANs to create more transparent connections with the Internet and to develop methods for maintaining the security and reliability that have been the key value-added features that have attracted users. VANs will be very cautious about risking their hard-won reputations for security by connecting to the public Internet. Stories are publicized frequently as another computer buff figures out how to break the code underlying current security and encryption techniques.

For NCES, one issue is the selection of the WAN to use. If indeed VAN-to-Internet connections become universal and functional, then NCES, as all other users, will be able to select the WAN or WANs that meet their needs the best. In the short term, the Internet's universality among public agencies and growing corps of proficient users argues strongly for its prominence in any planning.

Direct-Dial Connections

An alternative to these networks is a direct connection between two computers. A VAN or the Internet is not required to connect computers. The telephone companies provide connection using regular voice lines. One computer can dial another directly through their modems. This option provides for higher levels of security. Users can be required to have passwords for identification. Systems can also be set up to receive a call, then dial the caller back to ensure that your computer is really talking to the one identified as the caller. Direct-dial connections incur any applicable long-distance call charges. However, for the cost of a call, security can be significantly enhanced.

What are the implications for NCES of the ubiquitous accessibility of networks and the growing use of them by education-related agencies? The availability of universal network connections among NCES's trading partners nationwide provides tremendous potential and impetus

for changes in the way data are collected and reported. This is not a new realization for the agency. In fact electronic exchanges have already been implemented in several areas. What this paper is pointing out is that now is the time to make that full commitment to use of electronic networks. There should no longer be a hesitancy to move forward as soon as possible with conversion of NCES data collections from paper to electronic.

NCES sponsored 30 automation feasibility site visits to state education agencies from 1992 through 1994. During these visits, numerous examples of states' early attempts at using floppy disks for submitting reports were found. Both visits that included higher education interviews found disk reporting being tried. Reactions were universally positive, and plans were in place for expansion of the process.

Software and Applications

Relational Databases

Whether in physical reality or in concept, the emergence of relational databases has changed how NCES can plan for the future. A relational database stores data in the form of tables. They are powerful in that they impose few assumptions about how the user is going to want to access or analyze the data. Consequently, many individuals can benefit from the same database by using it in many different ways. In contrast, a flat-file database is self-contained in a single file. Everything a user needs must be in that same file to be used together. Relational databases are ideal for large information systems. They are also ideal for systems that will be used by many individuals with contrasting information and analysis needs.

This database issue is important, because the future design of NCES information system needs to take into account that all the data that will be needed may not, probably will not, reside in one location—or even within the NCES LAN. In line with this, NCES is very unlikely to define a file structure that will become universal across all the data systems that contribute to the NCES information system. In this context, the relational database design allows for the accessing of information across files for analysis.

Electronic Data Interchange

Moving data directly from one computer to another is called electronic data interchange (EDI). EDI is used by businesses for items such as purchase orders and invoices. Within the past 5 years, EDI applications have been developed for student transcripts and college loan applications. In fact, NCES was a sponsor of the development of the SPEEDE/ExPRESS standards for student transcripts. SPEEDE/ExPRESS is an approved standard by the American National Standards Institute (ANSI). Several vendors offer software to perform the EDI exchanges of transcripts. The Far West Lab in San Francisco provided copies of their ExPRESS.cal application for the Migrant Education Program pilot.

EDI is basic to moving NCES from a forms-based paper system to a data file-based, electronic system. Some states that have already begun submissions of reports from districts to their state education agencies on disks use a different technique. These processes involve filling out what

looks very much like the paper report forms on a spreadsheet or a word processing template, then making a copy to submit. EDI is the sending of a data record in a specific data format. The computer on each end of an EDI exchange can interpret the format and produce the types of reports on screens or paper that people are used to seeing.

Remember the last paper transcript you saw and compare that image to the format displayed in Figure 1.

This is an EDI record. The computer sending it and the computer reading it know exactly what each part means and how to interpret the contents. Each line is a "segment" containing information in one area. For example, the **SUM** line indicates 6 semester credits earned out of 6 attempted for all work taken at the sending school where 0 is the lowest possible grade average, 4 is the highest, 3.5 is the student's grade point average, and N means the grade point average cannot exceed 4. An entire transcript can be translated using these segments and their code tables.

You do not ever have to see this EDI language, because the computer translates everything into your local file format. When you see the information interpreted and printed as a transcript or displayed on a computer screen it looks no different than any other transcript.

In the absence of a national standard such as ANSI's SPEEDE/ExPRESS, commercial vendors would use their proprietary, and different, standards. Communications between vendors' systems would continue to be difficult.

Figure 1—Example of a SPEEDE/ExPRESS Electronic Record

ST*130874300021 N/L
BGN*00*87400021*900910*1530*ES N/L
ERP*DD*B48 N/L
REF*SY*123456789 N/L
DMG*D8*19790109*M*I*0*1US N/L
IND*US*FL N/L
N1*KR*Eastside Elementary School*77 *123456789101*9876 N/L
SUM*S*B*Y*6*6*6*0*4*3.5*N N/L
SES*198298*1**2*Fall Term*D8*19829824 N/L
SE*11*874300021 N/L

Productivity Software

Intelligent software applications that make work easier are emerging daily. The trend is for more of the work tasks performed to be automated. The benefits are not just for the worker who receives assistance with accuracy, finds the need to redo or recreate work less frequently, and is able to focus on more critical, clerical tasks. The benefits are also for the organization that receives data

on the processes of the business and the work that is being accomplished. As the worker performs duties, the software does the work of keeping the records and producing the reports.

For NCES, the implication is that automated software applications can be developed that perform the technical aspects of reporting, look for and alert the users to data quality issues, and reduce the burden for those providing the data as well as for the NCES staff receiving the data.

Voice, Video, and Text Processing

An examination of the NCES data collection forms reveals that much of the information reported is textual. Software is available now to analyze the content of text, to search for key words, and to index topics. Voice recognition technology has advanced to the point where it is practical to translate speech into text. Imagine a performance report for Title I compensatory programs containing a voice message describing program implementation issues. Video is becoming a more common method for recording program delivery levels. Video is being analyzed for communications patterns. A combination of video and voice recognition could be used to create a text record of classroom activity, then to produce a content analysis.

Practical use of these technologies does not appear to be possible within the short term. The issues of interpretation and use would overwhelm staff who are already challenged by the quantity of data being collected. However, future visions and plans should recognize the potential for these types of data collections and analyses.

SUMMARY OF IMPLICATIONS FOR NCES

What does this all imply for NCES? NCES should position itself to ride the wave of automation in the nation. The trends described here are as follows:

1) Faster computers will allow NCES to expand the amount of data collected, analyzed, and reported while potentially reducing the time and burden imposed on clients and NCES staff. The burden imposed by the quantity of data collected will decrease as an issue over time. Burden will be a consequence more of the availability of data versus the need to collect unavailable data. Of the data that are a part of an existing automated system, the burden to pass them along to another agency for analysis lessens as computers become faster in processing large databases.

2) Increased storage capacity on computers will allow NCES to collect and maintain as much data as is reasonable to collect based upon the information needs of clients. The amount of data to be collected will not need to be limited by the problem of where to put them when they are received.

3) The universality of networks will allow NCES to collect data electronically, communicate to clients electronically, and make available its analyses and reports electronically. Not only will virtually all agencies have access to networks, they will be wanting to use them. There will be a demand from reporting agencies that NCES accept all submissions electronically to avoid the burden of creating paper reports from local data files.

4) EDI standards and software will make electronic data exchanges over these networks efficient, effective, and affordable. EDI standards such as SPEEDE/ExPRESS may not become universal as the formats for maintaining data within agencies' databases. However, translations to EDI standards will become almost routine in order for agencies to exchange data files without rekeying information. In the short term, use of word processing templates and spreadsheets will begin the process of paperless reporting. NCES should continue to take an active role in the development of voluntary standards that facilitate electronic communications.

5) Relational database concepts will be applied to a distributed information system that will allow access to data across individual federal agencies' files. Where EDI standards provide a common language and process for exchange, database designs will allow for sharing or accessing of more complete data files by multiple agencies. For example, the Migrant Education Program in South Carolina envisions querying a data file in Georgia to locate the education records for arriving students. Then the Georgia schools will use SPEEDE/ExPRESS standards to send the students' records from their last school in Georgia to their new school in South Carolina.

6) Productivity software will automate information management tasks to the extent that staff will insist upon computer applications over any remaining manual processes. Software will continue to evolve to be more complex, more intelligent. Most of the tasks that do not require individual judgments will be handled by computers, with staff monitoring and intervening only when necessary.

7) NCES can achieve the benefits of an individual student-level database without the problems of creating a single one in NCES. The emerging networks and standards can create a national distributed information system. NCES would be able to query each state database to conduct analyses without having to maintain individual records centrally. The requirements for confidentiality can be maintained, and NCES would have access only to those data elements that are available to them by federal and state laws.

SPIN-OFF EFFECTS

The changes enabled by the advances in technology as described above do not come without their own spin-off effects. These are the indirect effects that occur as a consequence of a change.

Transfer of data processing responsibilities from a centralized data processing department/staff to the NCES staff or to the staff within other agencies is a major change. This transfer of responsibilities may also take the form of moving tasks from a few key staff members to a larger set of workers. As productivity software is installed, as networks make direct connections between agencies, as agency staff perform the actual data management tasks, the need for an external service group traditionally called the data processing department changes. This has benefits when staff are no longer waiting for their work to move up the priority list. Data are on your own computer, available when you need them. This has a downside when your staff must be retrained to perform new duties.

The role of the traditional data processing department shifts from one of actually doing the processing to one of supporting those who are. Programmers and systems professionals who are grounded in mainframe computer operations can have a difficult adjustment to the very different skills required in a distributed information systems environment. Data processing professionals will be called upon to support others and their applications.

The quantity of data will increase, especially as nonaggregated data are reported. More data and more analyses will put pressure upon staff to monitor and assure the quality of statistics and the reliability of analyses. Quality assurance procedures will need to be adjusted accordingly. Today, NCES calculates and issues official statistics on the nation's schools. With a data warehouse providing access to many researchers and interested organizations, almost anyone can calculate his or her own versions of those statistics. This would lead to a healthy debate as alternative analyses and perspectives are examined. This can also lead to the necessity for NCES to defend their formulas and calculations. Some form of quality check will be needed to respond to the alternative statistics offered by individuals and organizations. All of these will not follow the same rigorous standards NCES staff will follow when producing statistics.

IMPACT OF TECHNOLOGY CHANGES ON HUMAN RESOURCES

What businesses have discovered and learned to plan for is the impact of changes in technology on their people. Hardware and software costs are usually less than the associated costs for training and supporting the users. Within education organizations, the impact could be even greater. Staff development has historically received low priority—even for activities that are clearly directly related to the primary learning focus of the organization. Much less emphasis has been given to technology- or data-related issues.

Beyond retraining individuals and modifying hiring requirements and practices, organizations must restructure their staffing charts to reflect changes in the activities of staff. For example, state education agencies are already changing formerly secretarial positions into software applications support and training positions. As managers do more of their own word processing, there is less to type, and other traditional secretarial tasks also decline.

For NCES, planning must recognize the changes that will be imposed upon other agencies who must adjust to more automated processes. NCES will need to consider its role in retraining state and local staff. Development of training materials, sponsorship of workshops, and other support should be considered. NCES and its trading partners will be revising their job descriptions and the qualifications sought for new staff. Promotion and assignment decisions will reflect more of the technology-related skills necessary to implement and maintain the automated systems discussed here.

A VISION OF FUTURE AUTOMATED INFORMATION ACCESS

NCES will create a vision for future data collection and reporting. With the technology advances described in this paper, the following aspects of a vision seem reasonable:

- There will not be reports to fill out and submit. The concept of a report will change from being a document that someone fills out by collecting, calculating, and entering information. A report will become an analysis created from data sources available within an organization's information system.

- Most of the surveys and data collections that occur now will disappear. The concept of a survey or data collection as a specific request made for information on a report form will change. The individual needing data will go directly to a data file and read/copy what is desired. The concept that a survey or data collection occurs at a given time will shift. Data can be harvested from data files as needed, multiple times during a year. As an alternative, a reporting agency can upload (submit) their data as they become available.

- Almost all data about education will come directly from databases that are built as a natural part of conducting the business of an education agency. As more work is automated using productivity software, data documenting that work will be maintained as part of the software's task. Grades within automated grade books, records of transcripts sent to colleges, numbers of free meals served, and so on, will be recorded as these actions occur. When the data are needed, the data files will be read directly.

- When a new mandate for data collection and reporting arises, existing data sources will satisfy most of the requirements. New mandates for information will be checked against existing data sources. Only those elements that are not already available will need to be added to the information system.

- School personnel and education agency and staff will not think of the paperwork burden imposed by other agencies, because most of it will be transparent—accomplished as routine within their own automated management systems. Instead, considerable thought will be devoted to keeping information systems compatible, linked electronically, and current.

- Data will be collected and entered into these management systems because they are useful to the schools and education agencies. The best quality control is achieved when the persons responsible for the data depend upon the data for their own purposes. When the data have meaning, the individuals responsible for the data know when they are accurate and complete. Burden will not be a major issue, because the data are useful to those producing them.

- When the educators, news media, researchers, parents, and others have an information need, they will access data directly through an electronic network, in their own offices or homes, and create just the reports they need. The concept of huge volumes of statistical reports will change. The statistics will exist in data warehouses rather than on paper. Some statistics may not even exist until they are requested. Many more statistics can be produced than would be in a printed volume. Audiences can access statistics or in some cases the data used to calculate those statistics.

- Confidentiality will be maintained within the automated systems, allowing access to those with clearance and denying it to others. Directories and certification processes will determine an individual's access to data.

- A common data dictionary will define data elements and statistics along with the periodicity of their collection. Agencies will voluntarily use common data dictionary entries to ease the burden of translation when information is exchanged.

- Electronic networks will connect agencies, so data can be harvested from databases according to the periodicity specifications. Agencies will be able to read data directly from each other rather than having to make a request and await a reply. Data within each agency's information system will be categorized as public, restricted, or confidential to ensure that confidentiality rights are protected.

- The system will be voluntary, and compliance will be almost universal. Compliance will come from a common understanding of the benefits. Some entities will choose not to automate, and others will have local laws limiting participation.

- Cost savings will offset expenses, and the savings in personnel time will refocus resources on the primary mission of the educational agencies. Teachers will have more time to teach, librarians will have more time to manage their collections, financial aid officers will have more time to counsel students, etc.

- The components of the system will develop over time, joining together as they become available. Every agency will not participate from day 1. A paper system will be needed for some. Over time, the vision will become more universal.

- NCES will enable the system to develop by setting national standards and encouraging states to follow their example. The role of NCES will be key. As a facilitator of standards and a collector of data at the national level, NCES will be a model, a sponsor, and a participant.

- Reports will be printed by users as they are needed; many will be read on a monitor and no paper will be used. The concept of printing and disseminating a report will change. Most reports will be placed within the data warehouse and audiences will access the parts they need. Printing can occur at the reader's location rather than at the Government Printing Office. Printing would be at the reader's expense.

- The quality of education data will improve dramatically as use of the data motivates everyone toward accuracy, and the source of data becomes the management system that educators depend upon for their own work and productivity. As the data are used by more individuals and for more purposes, the benefits of accuracy and the risks of poor data increase.

- The ultimate purpose for collecting, analyzing, and reporting education data is to improve learning. With an open information system informing decision making, improvements in the quality of instruction and the management of education agencies will occur at a faster pace than ever before.

To achieve this vision, NCES will need to employ technology effectively. A major part of its planning must include a data warehouse or an alternative that achieves the same level of access to its information resources.

THE NCES DATA WAREHOUSE

Currently, access to NCES data and publications takes the form of printed documents distributed through a dissemination process involving mailing lists and orders through the Government Printing Office. Recently, some NCES publications have been placed on an Internet World Wide Web page for access. Access in the future should have many options from print to electronic files.

The technology trend and advances described in this paper support a direction already evident in NCES's planning—to develop a data warehouse. A data warehouse is simply a location where someone can access information electronically. As with many terms in the technology arena, there are differences in the characteristics people attribute to a data warehouse. A major attribute that varies across users of the concept is the level of aggregation for the data provided. To some a data warehouse is like a library containing books with statistics and analyses already accomplished and described. To others, a data warehouse contains an organization's raw data—available for analysis. For NCES, both are appropriate. With very few exceptions, NCES's data are public, as are any documents produced. Therefore, protecting the confidentiality of data or limiting the distribution of reports is seldom an issue.

NCES is on target with its current effort to build a user-friendly interface with its data warehouse. The key to widespread use for any computer application is utility and ease of use. NCES's concept is to give users the ability to search files for the data or other information they are seeking, then to download them as desired. The contrast with this and the current printing of large paper volumes called digests of education statistics is mainly with the ability of the user to find what is sought online rather than to find a printed volume and look up the statistics. An added bonus for users will be the ability to create tables and reports containing the information in which they are interested, rather than being limited to the manner in which data have been presented on the printed page.

The data warehouse can also function as a receiving point for data. Submissions by states can be uploaded to the data warehouse as soon as they are ready. This method can also be integrated with the harvesting concept. Both can operate within the information system.

Of course, the data warehouse concept should not stop with NCES. In fact, at least one state, Hawaii, has a functional data warehouse now, and others have them in the planning stage. The description that follows considers the benefits of a collection of data warehouses that are connected by networks and common EDI standards.

In this possible model, there would be multiple data warehouses containing in the aggregate all of the important and useful education data from across the nation. NCES would have one. Many individual states would have one each. Some states might join together to share a common data warehouse. Some states might use a commercial service. Within some states, there might be regional

centers that provide this function. Some districts may be large enough to justify operating their own. Even some schools, especially private schools, may want to establish their own. The fact is that the number and nature of the individual data warehouses and who is participating in each is not consequential. What is important is that they all use certain standards for EDI. They might also all use common database structures or formats to allow direct access to selected files by other organizations.

In the diagrams in Attachment A, NCES is shown as building and maintaining a central directory of agencies. This directory would build upon the Common Core of Data directory information currently collected. In addition to current data elements, this electronic directory would contain each agency's network address, contact persons, access information, and other usage parameters. The directory could be updatable directly by each agency. Thus, it would become a self-maintained directory.

The collection of data warehouses would be a distributed information system to the extent that common standards are used to store and access the contents of each. The contents accessible this way would be restricted to those data elements that each agency is authorized to provide to other agencies. This set of data is called the Confidential Data File. Contents would include items such as individual student and staff demographics, immunization data, course and grade data, assessment results, and program membership data.

A second data file within each data warehouse would be called the Public Data File. The contents of this file would be available to anyone. This would include such items as aggregate demographic statistics, enrollment statistics, financial data, assessment reports, and campus descriptions.

Behind these two files that are accessible to persons outside the agency would be the source data files. These source data files would be the master copies of data and would contain all data elements. These files would be secure, and users of the data would access copies of these files.

National education data and publications would reside in the NCES data warehouse. Communications between data warehouses or with individuals would be through the Internet, VANs, or direct dial as established by each agency.

How would the existence of these data warehouses affect NCES's data collection processes? Instead of sending out surveys to be completed or other forms-based data collections, NCES could connect to each data server for each data warehouse and download the information needed. The timing of these downloads would have to be known by all. Each data server should also contain an indicator of the status of the data for download by NCES. Each agency would be left an electronic receipt for their data.

In order for the data warehouse network to function, there must be national standards for data definitions and formulas. This is equivalent to a common data dictionary. However, even without a common data dictionary, participants in the distributed information system can communicate by translating their local data to a common standard such as SPEEDE/ExPRESS.

In Attachment B, the relationships among the levels of education agencies are described. The data within each level's information system are shown as being for internal use only, or as being shared with other levels. For either direct reading or harvesting of data to function, these relationships must be clarified and the data elements that fall within each category must be identified.

Timeframe for the Vision

Portions of the vision are in place now. Some states and some NCES activities are following, or more appropriately, leading the vision. The technology required for this vision to be fully implemented is already available. The hardware and network components are the most advanced. The productivity software will continue to be developed as agencies call for it to advance. It is reasonable for NCES to target converting all of its data collections to EDI by the year 2000. Activities may need to provide for paper submissions as an alternative for some.

The transition of NCES to automated data collection and a data warehouse is an ongoing, developmental process. There is not a turn-key system that can be purchased and installed.

Assumptions for Planning the Future Systems of NCES

The previous discussion of the advances and trends in technology points toward a set of assumptions that NCES should consider in planning its future information systems.

1) NCES can expand the amount of data collected, processed, and reported using faster computers. The time and burden imposed on clients and NCES staff will be less because of this processing efficiency.

2) NCES can collect and maintain as much data as is reasonable based upon the information needs of clients. Increased storage capacity on computers will allow reporting agencies and NCES to handle significantly larger data sets.

3) NCES can collect data electronically, communicate to clients electronically, and make available its analyses and reports electronically using national networks. Current forms-dependent data collection systems can be replaced with EDI-based systems with the expectation that reporting agencies can comply and participate.

4) Electronic data exchanges over these networks will be efficient, effective, and affordable. EDI standards and software will make these exchanges practical for agencies.

5) The change to EDI and other automated systems will require significant retraining of staff at all levels.

6) Allowing direct access to information in a data warehouse will increase the use of NCES information.

Process for Exchanging Data Among Education Agencies

National Center for Education Statistics Data Warehouse

This data warehouse functions to provide all audiences access to the data collected, the analyses conducted, and the publications produced by NCES. Electronic access through telecommunication networks provides immediate access without the necessity for NCES to print and mail copies.

THE CHALLENGE: To establish a process that connects schools, districts, states, and NCES for the electronic exchange of education data.

This challenge is great because records across schools, districts, and states exist in various formats and on a variety of computer systems. In addition, there is no funding available, even if agreement were to be achieved on the design, to build a single system. The ultimate solution must be an open-systems design that accommodates state, regional, local, and vendor solutions.

WORKING ASSUMPTIONS:

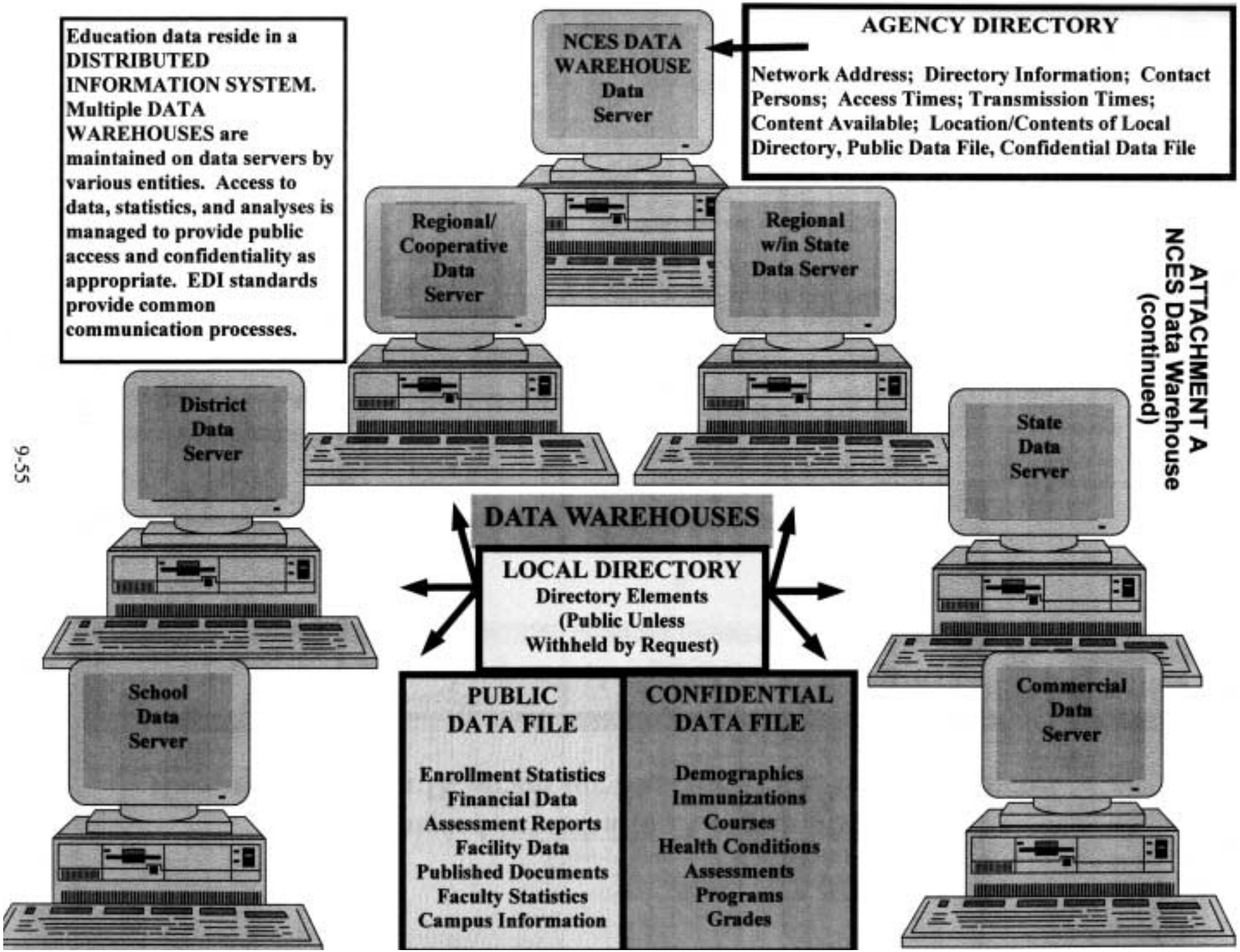
1. States will adopt different models to achieve the common goal of electronic data exchange.

2. Vendors will offer commercial solutions.

3. The role of the federal government will be to facilitate the development of standards and to use those standards for data collection. The federal government will not fund and build a system.

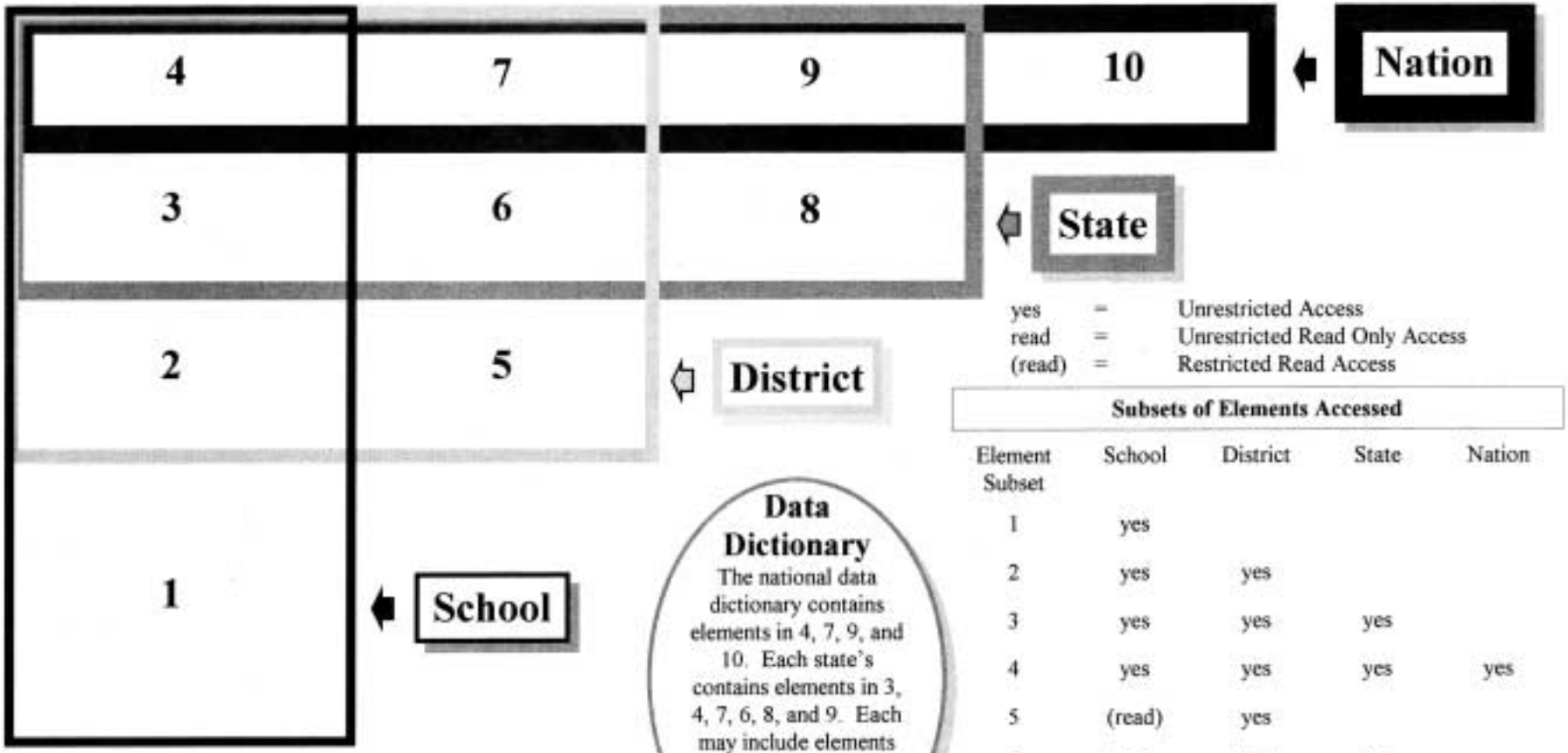
4. Participating schools, districts, and states will maintain control over access to their records.

5. Participating schools, districts, and states will determine the data they exchange.


Education data reside in a DISTRIBUTED INFORMATION SYSTEM. Multiple DATA WAREHOUSES are maintained on data servers by various entities. Access to data, statistics, and analyses is managed to provide public access and confidentiality as appropriate. EDI standards provide common communication processes.
AGENCY DIRECTORY
Network Address; Directory Information; Contact Persons; Access Times; Transmission Times; Content Available; Location/Contents of Local Directory, Public Data File, Confidential Data File
NCES DATA WAREHOUSE Data Server
Regional/ Cooperative Data Server
Regional w/in State Data Server
District Data Server
State Data Server
School Data Server
Commercial Data Server
DATA WAREHOUSES
LOCAL DIRECTORY
Directory Elements (Public Unless Withheld by Request)
PUBLIC DATA FILE
Enrollment Statistics
Financial Data
Assessment Reports
Facility Data
Published Documents
Faculty Statistics
Campus Information
CONFIDENTIAL DATA FILE
Demographics
Immunizations
Courses
Health Conditions
Assessments
Programs
Grades

ATTACHMENT B

Automated Information Coordination
Rectangular Venn Diagram for Sharing Data Across Agencies



yes = Unrestricted Access
read = Unrestricted Read Only Access
(read) = Restricted Read Access

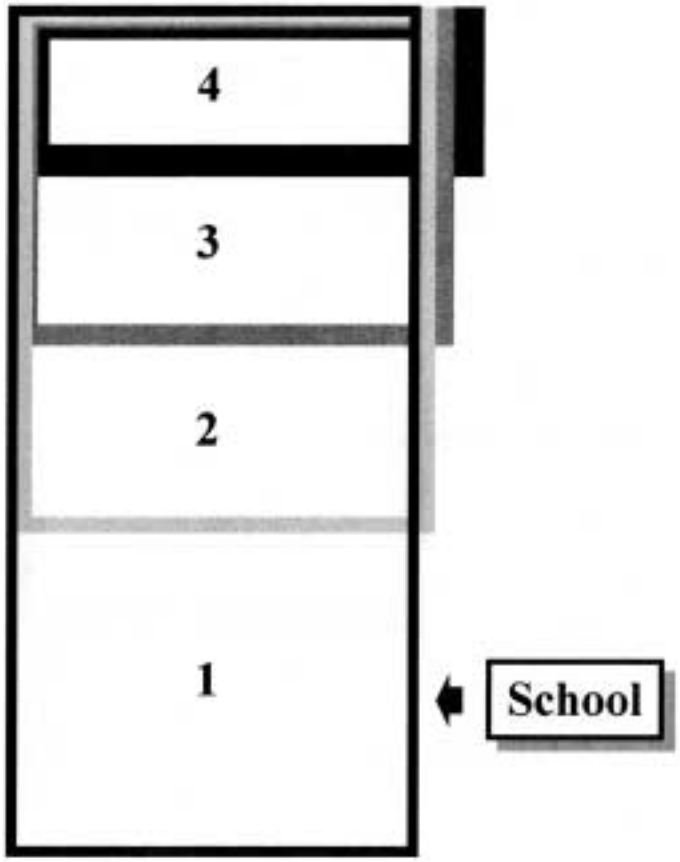
Subsets of Elements Accessed

Element Subset	School	District	State	Nation
1	yes			
2	yes	yes		
3	yes	yes	yes	
4	yes	yes	yes	yes
5	(read)	yes		
6	read	yes	yes	
7	read	yes	yes	yes
8	(read)	(read)	yes	
9	read	read	yes	yes
10	read	read	read	yes

Data Dictionary
The national data dictionary contains elements in 4, 7, 9, and 10. Each state's contains elements in 3, 4, 7, 6, 8, and 9. Each may include elements in other subsets to maintain comparability.

Example: A school maintains element subsets 1-4. Elements in 1 are used only by the school, whereas elements in 4 are shared by the school, district, state, and national agencies.

School-level data are characterized by being more individual to students, more anecdotal, more personal, and more instructional.



4 School-level data typically exchanged with national level:
- Gender, ethnicity membership counts
- School address, phone
- Grade levels served

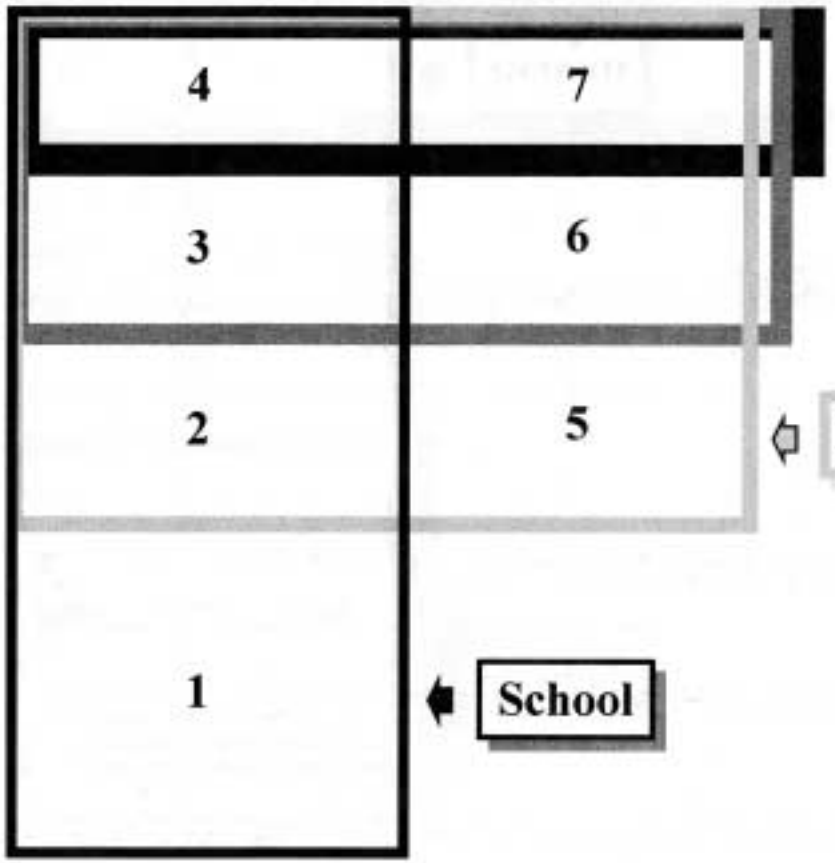
3 School-level data typically exchanged with state level:
- Attendance for funding formulas
- Gender, ethnicity membership counts
- Title I participant counts

2 School-level data typically exchanged with district level:
- Attendance
- Gender, ethnicity membership counts
- Teaching assignments

1 School-level data not typically exchanged with other levels:
- Book club orders and payments
- Homework and daily grades
- Emergency contact numbers and names

School data exchanges above the district level are typically performed by the district.

District-level data are characterized by more aggregated statistics.



7

District-level data typically exchanged with national level:
- Gender, ethnicity membership counts
- Title I participant counts
- District address, phone

6

District-level data typically exchanged with state level:
- Attendance for funding formulas
- Gender, ethnicity membership counts
- Teacher certification

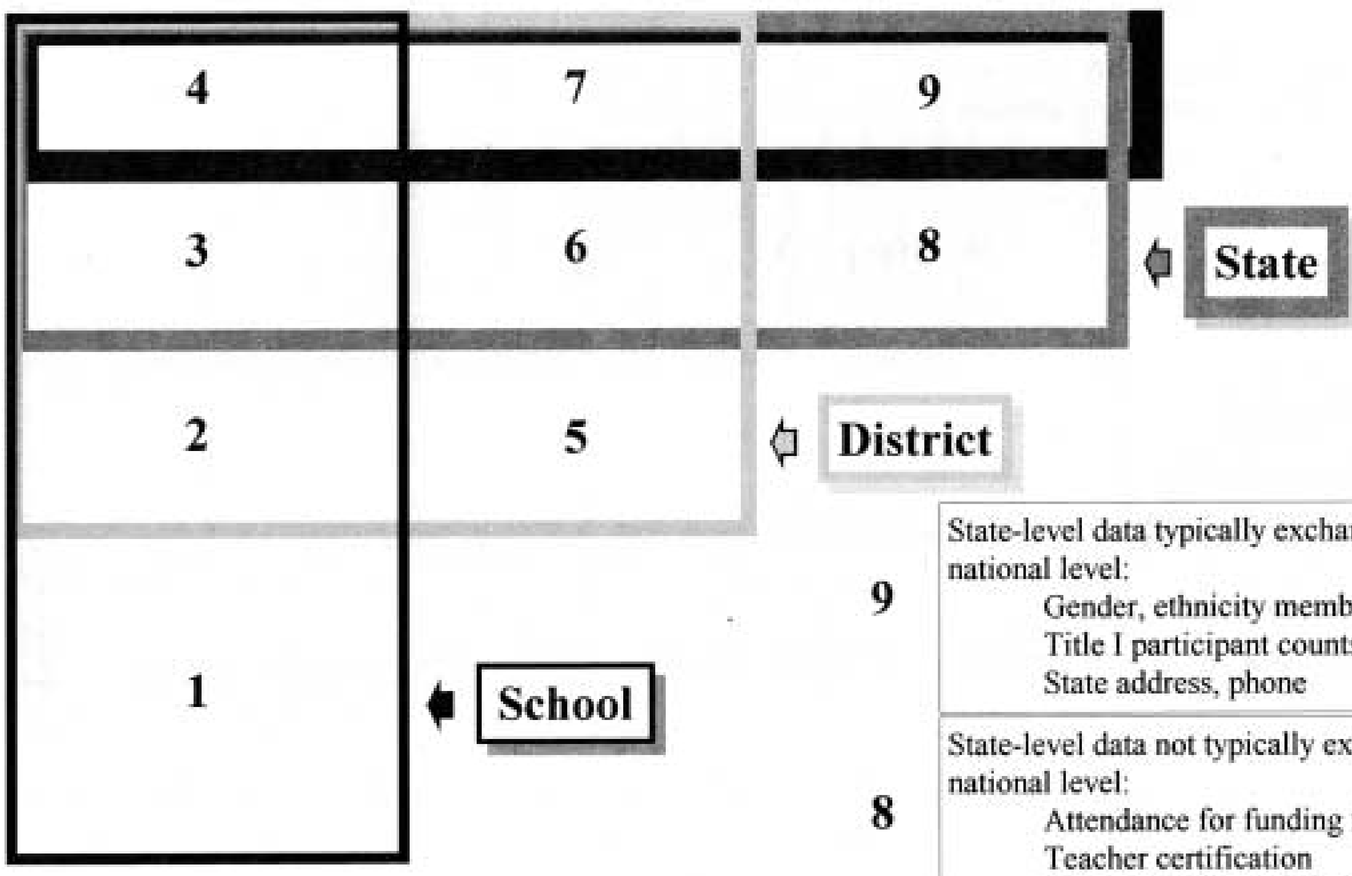
5

District-level data not typically exchanged with other levels:
- Statistics on test scores
- Attendance area statistics
- Committee membership

District data exchanges above the state level are often performed by the state.

ATTACHMENT B (continued)

State-level data are characterized by mostly aggregated statistics, except for states with student-level databases.

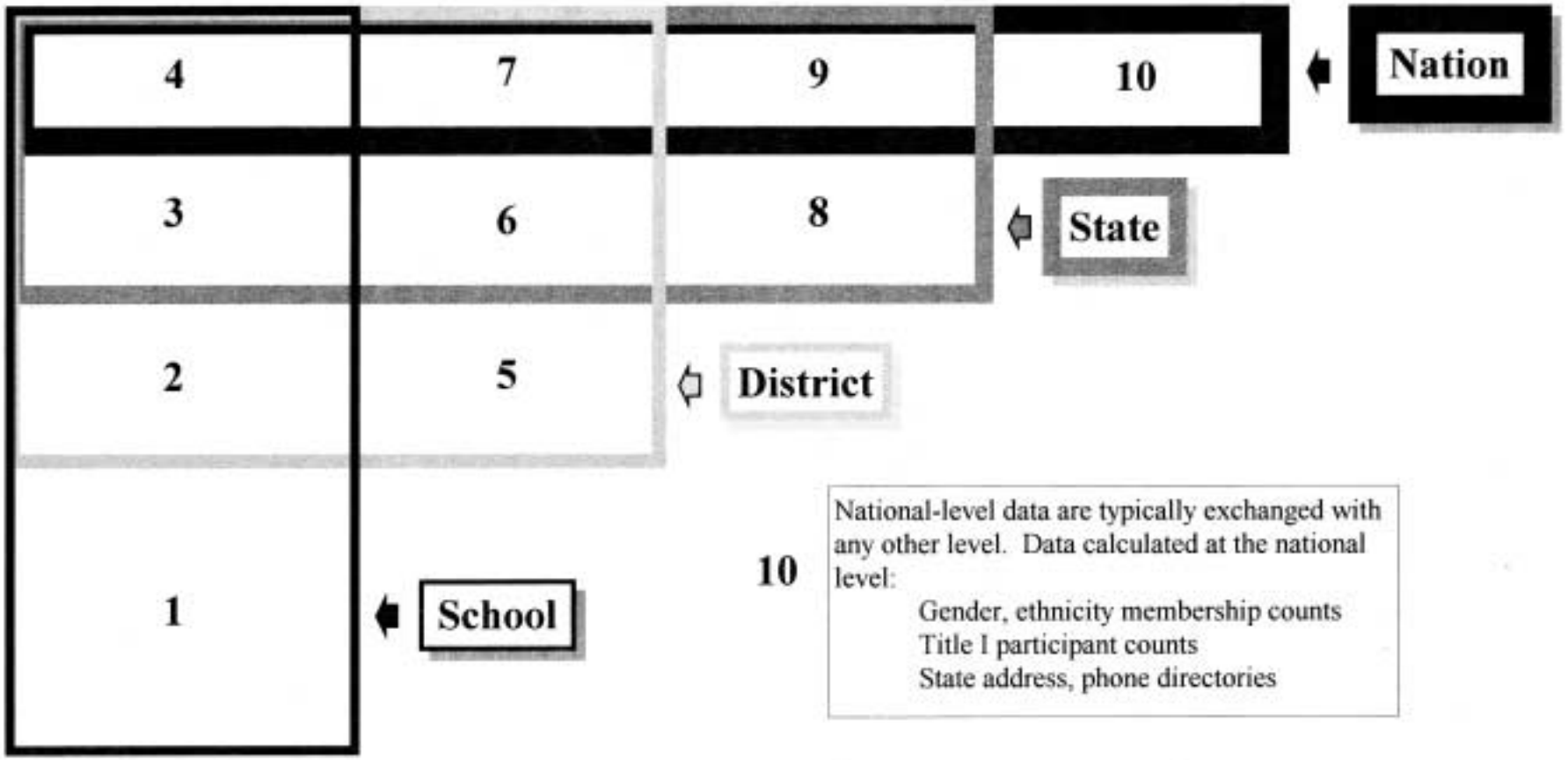


9 State-level data typically exchanged with national level:
- Gender, ethnicity membership counts
- Title I participant counts
- State address, phone

8 State-level data not typically exchanged with national level:
- Attendance for funding formulas
- Teacher certification
- State assessment statistics

ATTACHMENT B (continued)

National-level data are characterized by almost exclusively aggregated statistics.



Many Tenets of the Current Information Systems Will Change

The periodicity of collecting data will evolve from dates when forms are available and completion begins, to timeframes when files are built and extracts occur. Due dates will become extract dates, representing when data files will be read.

Data burden will shift from a term depicting a situation where work stops to document and report, to one where work is automated for efficiency and data for reporting are a byproduct of work activities. Data burden may become a term representing the overwhelming amount of information to read and interpret that is available on a given topic.

The producers of data, the entities being monitored or reported, will not have complete control over the information produced and known about them. Access to raw data will delegate to others the ability to generate statistics not ever seen by the target of the analysis. This loss of control by educators at all levels could slow the progress toward fully automated information systems unless groundwork is laid along the way.

Physical separations will be established between systems. Data warehouses will be created to hold the data that are accessible by a credentialed set of users, but the original management information of an organization will be more tightly controlled on isolated computer systems with fewer if any access links outside the physical facility of the organization. This means that levels of access will be established within this firewalled city of information. The sanctum sanctorum of an organization's data will be the original secure source files containing the most detail and most confidential elements. Confidential working extracts will be created and posted for certified users with a need to know. Multiple extracts will be created with the set of data authorized for a set of users. These extracts will be loaded onto separate servers without access beyond the organization and with only one-way access from the secure files. Limited access extract files will then be created, again directly from the secure source files. These files will be available on more universally accessible file servers, access to which is allowed for certified users from certified locations. The fourth and final set of extract files will contain public information placed upon a public access server in a read-only mode.

OUR CONCEPT OF DATA WILL EXPAND

Our concept of data has begun to expand as storage and processing capacities increase. Images, video, and voice have all claimed places within our automated data systems. How we will analyze and use them is expanding rapidly already. Content analyses, image scanning, voice-to-text translators, image-to-text translators are all becoming more sophisticated and allow analysis of now seemingly insurmountable amounts of data within a reasonable time in an automated fashion. Classroom observations recorded on videotape for example could be analyzed and coded with software programs designed to detect who is speaking, topics, movement, involvement of individuals, and even performance feedback actions.

CRITERIA FOR JUDGING THE FUTURE SYSTEM

In December 1985, Hall, Jaeger, Kearney, and Wiley prepared a report entitled *Alternatives for a National Data System on Elementary and Secondary Education*. Within that report, they proposed a set of criteria to be used for judging a national education information system. These criteria are revisited here along with a comment related to the characteristics of a system that would be consistent with the vision described above.

1) Comprehensiveness—the system must have a database capable of providing information on all pertinent aspects of elementary and secondary schooling, including the school setting, the schooling process itself, and the outcomes of schooling.

   Because the vision foresees inclusion of all data that are produced as a product of the conduct of the regular business of education agencies, it should be comprehensive. A limitation would be that data on the school setting and the schooling process would be available only to the extent that automated systems are used which would relate to school setting and processes. Is it the role of NCES to collect and process data? Certainly the vision sees an information system that includes a much more comprehensive database than is currently compiled.

2) Integration—the elements, files, and records in the database must be linked; all data sets must be capable of being related to one another.

   A relational database design would facilitate this. A common data dictionary would be necessary along with definitions and formulas for all calculated statistics. This criterion would require that links already exist at the local, district, and state levels.

3) Micro Record Format—all data must be collected and stored in micro record format, with a micro record being defined as a datum on an individual person or an individual entity.

   This is problematic at the national level. Despite the increased storage capacity and speeds of computers, a data file of over 50 million student records would be cumbersome. This criterion can be met if we accept the idea that there will be such a micro record for every individual at some level of the distributed information system. Some individuals records will reside only in a school's database, others at a district level. A few states have individual records systems for students; many more have them for staff. There is no mandate or plan for NCES to collect personally identifiable records at the national level.

   However, with the ability to harvest data or conduct analyses on data distributed across multiple data servers, the functional intent of this criterion could be met.

4) Representativeness—in addition to being nationally representative, the information in the database must be representative of each of the 50 states, as well as representative of other important variables such as sex, racial-ethnic composition, urbanization, and so on.

   This criterion would be met with the participation of all states. The content collected by current Common Core of Data surveys addresses the intent of this criterion. The

collection and storage of disaggregated data, individual student records, and generally more detailed data provide the opportunity for post hoc analyses that consider additional variables.

5) Accuracy—all data must be verifiably accurate; they must be subjected to rigorous quality control procedures including audits, reinterviews as a routine part of data collection, controls on data entry and data processing, consistency and completeness edits, and regular and routine calculations of measures of variance.

   The emergence of large, central databases into which data are reported for state information systems has popularized the term "desk audit" to represent quality checks that are performed on data that have been reported. An individual sitting at a personal computer on a desktop can run verification and audit software. These checks look for data out of normal ranges, illegal codes, missing data, etc. NCES or independent groups would be able to perform reasonableness audits or even follow up with source rechecks if data are provided in a data warehouse.

   Productivity software used would contain validation checks as data are entered (the level where errors are most commonly created and where they are most easily resolved). EDI software contains validation checks for data sent and received. However, errors that fall within the normal range of data can typically be found only by the provider of the data.

6) Comparability—data from different jurisdictions must reflect the same concepts and definitions; common units of reporting and common definitions are necessary precursors of useful data aggregations.

   NCES has traditionally provided clear definitions and formulas for the aggregated statistics it collects. How faithfully the data providers follow these standards varies. Automating the collections will not solve any current problems. However, adoptions of common data dictionaries and use of common software applications can emphasize the definitions that are to be followed.

7) Timeliness—in general, data must be limited to that which can be collected, stored, and analyzed within three months and reported to policy makers within the year.

   The vision uses technology to address this criterion directly. Taking data from existing systems, electronically exchanging them, and providing the capability for faster analyses using large data sets all contribute to timeliness. Reporting to policy makers can be improved with electronic availability from the data warehouse. One concept with the data warehouse is that states could post their data as soon as it is available, then they would be accessible by others immediately. The existence of a data warehouse can shift the burden to the users to know when data are available.

8) Privacy and Security—because some of the elements, records, and files contain information about individuals (e.g., personal identifiers necessary for longitudinal studies), strict confidentiality and security measures must be in force.

   Confidentiality and security challenges exist with paper systems. They receive greater attention with automated systems because of some highly publicized events and the very

real risk of hackers. Electronic systems allow for very elaborate security processes. Even these are not failproof. However, the required sophistication of a successful hacker can be pushed higher and higher, and automated systems can document access. Many individuals believe that the security of electronic systems is superior to that of paper systems.

9) Processing and Analysis—a specific schema must be available for processing the micro records in a manner designed to optimize the analytic capacity of the system.

   The increased speed and storage capacities of computers contribute to this. The ability to analyze the larger data sets has improved considerably since 1985. Analysis software and the emergence of relational databases have boosted the capacity of researchers to perform analyses.

10) Information Flows—the system must be capable of screening and matching its reports to meet the particular needs of users; a wide array of reporting formats and access mechanisms must be available to serve the different users; specific priorities must be set for meeting the different timelines imposed by the needs of the users.

    This is an excellent example of a criterion that is well served by technology. With a data warehouse, users will be able to search indexes as well as text to find information matching their needs. Reporting formats increase with the addition of screen views, downloads, and user queries to produce just the statistics desired. With electronic access, users can get the information they need when they need it. The only constraint is that the data must be collected and already captured by the information system.

11) Costs of Transmission/Access—a pattern of shared user costs should characterize the system; rather than rely exclusively on federal support for transmitting information to users and/or providing them access to information, a national educational data system should also draw from a program of user fees and thereby increase its capacity to serve the differing needs of its users; equally important, transmission/access modes should incorporate the latest developments in electronic communications technology.

    The user pays the cost when connected through an electronic network. Whoever is connecting pays the transmission fees. The costs for establishing and maintaining the data warehouse would not be easily shared with the users. The cost of that type of billing might exceed the actual fees recovered.

The conclusion of the authors was that the only criterion met by the NCES system of 1985 was Privacy and Security. Interestingly, this is the one that could be the most controversial with an electronic system. For the other ten criteria, an automated system using a data warehouse concept has the potential for significant improvements.

CONCLUSION

Ensuring that NCES's data collection, storage, analysis, and reporting processes take full advantage of technology will be a process, not an event. This transition will require considerable training and support for both NCES staff and the staff of its data providers. When evaluated against the criteria described in 1985, the vision of the future as described here would be a significant improvement over past and current systems.

Discussant Comments

BARBARA S. CLEMENTS

These comments address issues raised in two papers: *Administrative Record Opportunities* by Fritz Scheuren and *New Developments in Technology: Implications for Collecting, Storing, Retrieving, and Disseminating National Data for Education* by Glynn Ligon. Both papers describe important issues that must be considered by the National Center for Education Statistics (NCES) as it seeks to make its data collection activities more efficient and as it responds to technology changes occurring in the sites where the data originate. In these comments, I provide some background comments, and then react to the papers from two perspectives: the user perspective and the provider perspective.

Administrative records exist in all schools, districts, and state education agencies in a vast array of formats and with a variety of contents. While many schools, districts, and state education agencies may have some data automated, most are still heavily reliant upon paper records. Two examples illustrate this point.

About 10 years ago, when Texas was implementing a Career Ladder, a teacher from a tiny district called to see about getting evaluated for the Career Ladder. In the course of the conversation, she was asked where her personnel records were kept. She thought for a minute, and then said that she believed they were in a shoe box under her bed. Eight years ago, when I moved to Washington, D.C., I went to my son's school to get a copy of his high school transcript. I was given a photocopy of a paper document that had computer labels pasted on it. It was obvious that some parts of his student record were computerized, but the paper document was still used to compile his course data. According to my school contacts, these two examples illustrate the lack of technological sophistication with administrative records that still exists today at the school and school district level. I have heard of very few places in elementary and secondary education where there is a fully automated administrative records system that can handle the types of electronic exchanges and sophisticated analyses that are technologically possible today.

How data are used at the local and state levels is important when considering data quality. My sense is that in most schools and districts, most data are recorded because someone thinks they should or because someone requires it, such as the state legislature or the federal government. Few state or local education agency staff members have the time or opportunity to think about how data can be used to assist in providing quality instruction to children, the primary goal of the education system. Since the data have “little utility,” there is no impetus to ensure comparability or timely updating. If NCES is to get useful data from state and local administrative records, it must develop ways to encourage and help data providers to collect and provide comparable, complete, and timely data.

Data User Perspective

As a data user, I have several comments about the papers. The Ligon paper describes the design for an automated administrative records system that can provide data access and give flexibility for data analysis to all levels of the education system. The Scheuren paper describes what valuable information is available when administrative records can be collected. Timely data availability is an important benefit both authors describe, and it relates to the ease with which electronic administrative data can be transmitted to different levels of the education system.

Current lag time in getting data from NCES from the Common Core of Data and other surveys has been frustrating for many data users. The work that can be done by NCES to streamline data editing routines and speed up reporting and data tape availability is essential. An electronic system such as the Ligon paper describes can allow data to be submitted from original sources with no rekeying needed; thus, the errors in the data should be minimized, and this should speed up the process of making data available to users. Such a system requires preparation at all levels of the data system; therefore, it is important for NCES to be ready to accept electronic data and process them quickly and efficiently.

Both papers indicate that moving to electronic submission of administrative records can provide more comprehensive sets of data with which to work at NCES. Each time NCES asks for new data elements to be added to paper survey documents, there are state education agency staff members who complain about the burden of adding those data elements to their own collections and the lag time that is needed to get data from all sources. If states have access to electronic administrative records, it should be easier for them to get additional data elements if deemed necessary and provide them to NCES. This would make the data sets more complete and better able to respond to both policy questions that arise in Washington, D.C. and to questions asked by other NCES data users. This is another good reason for NCES to continue working with state and local education agencies to design automated administrative records systems with electronic transmission.

To me, the most important thing that should be stressed in the discussion about administrative records is the need for comparability in what is collected and provided to the different levels of the education system. NCES has been working for years with state and local education agency staff to build a consensus on how the data should be collected and reported to ensure comparability. This is stressed in the Ligon paper, but not in the Scheuren paper. Although all of the data maintained in administrative records at all levels of the education system need not be exactly the same, the portions that are reported up from the lowest levels must be comparable, or at least able to be crosswalked, in order for the data to be useful. Therefore, as a user, I believe it is important for NCES to continue efforts to promote comparability and standardization of those data elements that are essential for national data collection.

The Scheuren paper suggests that administrative data be used to track changes over time. I believe there is a real need to look at changes in student population, effects of participation in programs based on new federal or state policies, and other educational issues that can help decision makers in planning for school improvement. Besides tracking changes, NCES needs to explore ways of identifying effective programs through regular data collection activities, so that case studies or

further research can be done, not perhaps by NCES, but by others within the Department of Education, such as the OERI institutes.

Data Provider Perspective

There are several comments I would like to make from a data provider perspective. The work that NCES has supported related to providing tools to make the collection and transmission of administrative records easier are to be applauded. Burden is one of the most frequent complaints of state and local education agencies. State and local education agencies are looking for models of electronic data sharing that would be relatively easy to implement in technologically unsophisticated sites, and particularly ones that take into consideration existing equipment and planning for a system that can be implemented over time as funds become available. Such models would help state and local education agencies reduce their reporting burden and move toward providing more timely data. NCES has done some work to provide models for how data can be maintained, transmitted electronically, and used more effectively. The work NCES sponsors on confidentiality is extraordinarily important for all levels of the system. These activities have a great potential for payoff, and should continue.

Several areas still need the attention of NCES. First, NCES should look at all of the areas in its surveys where administrative records could provide essential data such as years of teaching experience, age, and so on, and plan to collect data in this way from schools, districts, or state education agencies to reduce the individual burden of individuals such as teachers who complete the surveys. To help promote comparability, stress should be placed on standardizing those data elements that will help data providers adjust their systems (or purchase appropriate systems) to meet future data reporting needs. As my data provider friends say, "Just tell me what you want and how you want it, and we will make it happen."

Second, many data providers need help with training on how to collect, report, and use data. At present, NCES provides a valuable service through the Fellows Program. Many state and local data providers would appreciate having models for how data can be presented more effectively for decision makers. For instance, videotapes are considered extremely useful by data providers because they can go back and review them when needed. Moreover, state data providers need help in training data providers from the local levels. Training is essential to getting comparable, complete, and timely data. NCES should place an even stronger focus on what they can do in this area.

And, finally, NCES should lead discussions with the health and human service areas about data sharing for the benefit of students. In education, we are constrained (and helped) by the Family Rights and Privacy Act (FERPA), and the other areas also have their professional ethics or other types of restrictions on usage. Currently, an important trend is on providing services to students through the schools. We are also encouraging teachers to make better use of student data when planning learning activities. NCES can play an essential role in looking at ways to reduce the redundancy in data collection and ensure that the data collected meet the needs of multiple users. NCES has worked with other units within the Department of Education, but now they should reach beyond the education boundaries. Data providers will greatly appreciate any assistance that NCES

can provide in convening and urging agreement on data formats and in considering ways that data can be legally shared with health and human services.

NCES can serve the education community well by keeping a focus on the future and what must be done to ensure that data collection efforts take advantage of electronic advances and meet future information needs.

DENNIS CARROLL

Fritz Scheuren's paper describes several opportunities for NCES. He broadly and boldly develops major implications for operations, staffing, and technology. Whether his predictions are realized within the next 10 years or not, NCES should prepare for the next revolution in analysis. This revolution is not statistical technique, but rather the predominance of administrative records as the birthing agent for data sets.

The paper rightly suggests that the quantity of administrative record data that may be tapped by NCES will continue to increase. Further, with faster, cheaper, and better connected computing, administrative records will be easier to use. Scheuren suggests that eventually data collections may become supplements for administrative data rather than the currently reversed situation. However, Scheuren failed to note the impact of restrictive privacy legislation, state budget declines, reinvention, and other political factors that are increasingly restricting access to systems of administrative records.

If Scheuren's notions are attempted, NCES must consider how far on the leading edge of this technological adventure it should venture. With limited budgets, NCES needs the administrative data to enhance limited data collections. However, with a shrinking staff and an apolitical mission, it is difficult to meet the demands of leading-edge status. The paper would be improved if it included suggestions about the areas NCES should try initially.

With an increase in administrative record quantity, there will be a compatibility potential that is limited by comparability. Imputations, as suggested in the paper, will become more prevalent. Without significant advances in imputation technology, the notions of fully or partially imputed data sets will be limited. Currently, it is doubtful that a little reported data can be appropriately combined with a lot of imputed data for meaningful analyses. For example, although imputation makes a constructed NPSAS possible with Central Processing System and IPEDS data as a source, the policy community probably would not use it.

Just as instrument nonresponse plagues survey collections, partial access will trouble administrative records. Biases associated with instrument nonresponse rarely have the impact of restrictions on access to administrative records. Analysts with access hold an advantage over those using the biased, even if fully imputed, data. How NCES should deal with this conflict is an important issue.

Finally, this paper rightly suggests that getting distributions “correct” should be more important to NCES than cleaning data case by case and variable by variable. Well-behaved data that adequately reflect the proper distribution(s) are simply better. Error estimation, modeling, and simple

statistics (graphical displays) feel better when using well-distributed data. In this area, administrative records can help, and they can help immediately. Many distributions can be known based on administrative records, without access to the microdata.

WILLIAM H. FREUND

Glynn Ligon was given the impossible task of describing "new developments in technology that have affected or will affect the collection and reporting of education data." This represented a difficult assignment at best and was impractical in this era of highly evolving telecommunications and eventual saturation of computers into our work and home environments. The issues are not technological changes—we know these will occur. Since these changes, particularly in telecommunications, will open up new markets for education statistics, the more important questions for NCES include the following:

- Who will be the customers of national education data?
- What questions will they ask?
- How should information be presented and retrieved?

It is important to note that these three questions do not even address the mechanics of technology (hardware and software). We will have the technology; the only issue is the extent of access within the education community and our customer base to this technology. Access is an important question for schools and districts without the financial resources to obtain high-speed Internet connections.

However, assuming access, just exactly how would these technologies affect the Center's data collection and dissemination of administrative records survey data? And is the Center doing anything now to take advantage of what is available?

Data collections for administrative records

Many people think that NCES continues to rely on paper forms for much of its data collection/survey work. Currently, the Center uses at least five different modes to obtain information from state agencies and colleges and universities. These include DBF files, ASCII-based data (on diskettes or tape), File Transfer Protocol (FTP), mail, and Electronic Data Interchange (EDI). However, only in library collections have we moved beyond these five somewhat traditional modes into an electronic forms mode. Only our library programs have turned in this direction, but plans are now under way to move more actively into electronic forms. At present, there are many "software" models available to guide our developmental efforts, specifically packages such as TurboTax™. These packages provide forms, year-to-year comparisons, and internal editing capabilities for consistency of responses.

But the important thing to remember is the impact of shifting to new collection practices. Technology will force data owners and providers to assume more responsibility for data quality and

timeliness. Thus, NCES's responsibility will shift toward developing and providing data owners with new and better tools to improve quality and timeliness.

Dissemination of Administrative Records

As with data collection activities, there is a misperception about how NCES disseminates its products. Computer tapes are no longer our primary mode of dissemination. In fact, we prefer *not* to send tapes. However, we are awash in new forms of products, including diskettes, CD-ROMs, tabulation packages (the Data Analysis System), Electronic Codebooks (ECBs), printed reports, gopher servers, phone orders, and, yes, a few tapes. In fact, these new products are invaluable to our customers. For example, the DAS software developed by Dennis Carroll and Larry Bobbitt obviates the need for users to understand complex samples, since the software handles the appropriate calculations for variances.

New techniques or methods are coming. For example, we are developing a World Wide Web (WWW) home page. We are also setting up an early release program for administrative records. And we are improving customer service in other ways, including expanding of the National Data Resource Center (NDRC). The NDRC provides tabulation services to customers without access to computers and/or appropriate software packages. But our real future in dissemination is embodied by our current initiatives with Structured Query Language (SQL) server and data warehouses.

Envision sitting in front of your personal computer; loading Excel onto your desktop; clicking on external data; linking to NCES via Internet; selecting data files of your choice; subsetting the file based on your own criteria; tagging those data elements that you want; and then retrieving the data back into your Excel spreadsheet. That scenario will be the ultimate dissemination program—providing the user with the right information, in the right form, in the right place, and at the right time. That scenario is actually viable today and is being tested internally within NCES and externally via point-to-point protocol.

Glynn Ligon's paper hits home on a variety of issues before these scenarios become a practical reality. First, you must be very familiar with file structures to use SQL server—user friendliness is not a design feature when it comes to data. Second, the user must have excellent documentation to use the files effectively. Electronic codebooks and DAS CD-ROMs are a step in that direction. But we should convert them to Windows so that users will simply press the F1 help key to obtain full descriptions of variable definitions and values. Another issue is for NCES to fully understand its customer capabilities. We might, as suggested by Fritz Scheuren, use the Common Core of Data (CCD) and Integrated Postsecondary Education Data System (IPEDS) to periodically survey our respondents and customers. We would then have some answers to the questions raised at the beginning of this commentary.

But easier data collections and expanded user access to data raise additional areas for the Center to consider and act upon. For example, standards and data comparability among survey respondents will become increasingly important. This is true across all levels of education, and NCES is currently promoting comparability via its efforts with the Cooperatives, handbooks, and EDI standards. We

also have to promote more leveraging of software if survey respondents are to make effective use of new technologies. While the cooperatives can play a role in this effort, responsibility will fall upon the states themselves. Finally, NCES must help users DIRTFT—Do It Right The First Time. In this case, “It” means drawing valid conclusions or findings from the various NCES data files.

With all these activities under way, NCES is addressing the challenges imposed by new technologies. I wonder what form those challenges will assume 5 years from now?

A Appendix A
About the Contributors

THIS PAGE INTENTIONALLY LEFT BLANK

About the Contributors

AUTHORS

Robert F. Boruch is University Chair Professor in the Graduate School of Education and the Statistics Department of the Wharton School at the University of Pennsylvania. A Fellow of the American Statistical Association, he has received awards for his work on research methods and policy from the American Educational Research Association, the American Evaluation Association, and the Policy Studies Association. Boruch is the author of nearly 150 scholarly papers and the author or editor of a dozen books on topics ranging from evaluation of AIDS prevention programs and social experiments to assuring confidentiality of data in social research.

David W. Breneman is University Professor and Dean of the Curry School of Education at the University of Virginia. He was Visiting Professor at the Harvard Graduate School of Education from 1990 to 1995, where he taught graduate courses on the economics and financing of higher education, on liberal arts colleges, and on the college presidency. As a Visiting Fellow at the Brookings Institution, he conducted research for a book entitled *Liberal Arts Colleges: Thriving, Surviving, or Endangered?*, published by Brookings in 1994. From 1983 to 1989, Breneman served as president of Kalamazoo College, a liberal arts college in Michigan. Prior to that, he was a Senior Fellow at Brookings, specializing in the economics of higher education and public policy toward education. From 1972 to 1975, he was Staff Director of the National Board on Graduate Education at the National Academy of Sciences, where he wrote on policy issues confronting graduate education. In addition, Breneman served as Executive Editor of *Change*, the magazine of higher learning. Dr. Breneman received his bachelor's degree in Philosophy from the University of Colorado and his Ph.D. in Economics from the University of California at Berkeley, and taught at Amherst College before moving to Washington in 1972.

Dominic Brewer is a Labor Economist specializing in the economics of education and education finance. His research has focused on educational productivity and teacher incentives, using large national databases such as High School and Beyond and the National Education Longitudinal Study of 1988. Examples of this work include an analysis of the effects of teacher education and quality on student achievement gains, the interaction between student and teacher race, gender and ethnicity, the effects of ability grouping on student achievement, and the effects of administrative resources on student performance. He has published numerous articles in academic journals such as *Review of Economics and Statistics, Journal of Labor Economics*, *Industrial and Labor Relations Review*, and *Economics of Education Review*, as well as other publications such as *Phi Delta Kappan.* Dr. Brewer received a Ph.D. in Labor Economics from Cornell in 1994, and holds a bachelor's degree

from Oxford University. He has been an Associate Economist at RAND since 1994 and is also a Visiting Assistant Professor of Economics at the University of California, Los Angeles.

Peter Cappelli is Co-Director of the National Center on the Educational Quality of the Workforce (EQW) at the University of Pennsylvania.

Christopher T. Cross is President of the Council for Basic Education (CBE) as well as President of the Maryland State Board of Education. Before joining CBE, Mr. Cross served as Director of the Education Initiative for The Business Roundtable and as Assistant Secretary of Education Research and Improvement (OERI) in the U.S. Department of Education. At OERI, he was responsible for the research, statistical, and improvement programs of the Department of Education. He joined the federal government for the first time in 1969 with the U.S. Department of Health, Education, and Welfare, where he served as Deputy Assistant Secretary for Legislation. From 1973 to 1978, Mr. Cross served as the Senior Education Consultant and Republican Staff Director of the Committee on Education and Labor, U.S. House of Representatives. Mr. Cross has written extensively in the education and public policy areas, and his articles have appeared in numerous scholarly and technical publications. Mr. Cross earned a bachelor's degree from Whittier College and a master's degree in Government from California State University in Los Angeles.

Fred J. Galloway is the Director of Federal Policy Analysis at the American Council on Education (ACE). In this position, he represents the interests of the higher education community before the executive and legislative branches of the federal government and is responsible for analyzing the effects of legislation on colleges and universities. Before joining ACE, Dr. Galloway was a member of the faculty of the Economics Department at San Diego State University and of the School of Business at the University of San Diego. Dr. Galloway received both a bachelor's and master's degree from the University of California at San Diego, and a doctorate from Harvard University.

Gary Hoachlander is President of MPR Associates, Inc., a consulting firm specializing in management, planning, and research for a variety of public and private clients. A nationally known expert on vocational education and preparation for work, he also serves as MPR Associates' site director for the work performed by the firm for the National Center for Research in Vocational Education at the University of California at Berkeley. He has conducted research and published on a wide variety of issues including industry-based curriculum, industry skill standards, performance measures and assessment, finance, and national education data systems. Dr. Hoachlander received his Ph.D. in City and Regional Planning from the University of California, Berkeley. He also holds a master's degree in City Planning from U.C. Berkeley, and earned his bachelor's degree from Princeton University, where he attended the Woodrow Wilson School for Public and International Affairs.

John F. Jennings is the Director of the Center on National Education Policy. The Center's purpose is to inform the general public, educators, and policymakers of the developments in school reform across the country and also of the changes in federal education programs. From 1967 to 1994, Mr. Jennings worked in the area of federal aid to education for the U.S. Congress. In that capacity, he was involved for the last 25 years in nearly every major education debate held at the national level as well as the reauthorizations of the major federal education programs including the Elementary and Secondary Education Act, the Vocational Education Act, the School Lunch Act, the Individuals with Disabilities Education Act, and the Higher Education Act. Mr. Jennings has also edited several books, published numerous articles, and writes a national newsletter.

Glynn D. Ligon is President of Evaluation Software Publishing, Incorporated.

David R. Mandel is Vice President for Policy Development at the National Board for Professional Teaching Standards in Washington, D.C., where he has primary responsibility for overseeing the Board's standards development efforts and education policy and reform program. Previously, Mr. Mandel was Associate Director of the Carnegie Forum on Education and the Economy; a Senior Policy Analyst in the Office of the Under Secretary of Education; and the National Institute of Education's Assistant Director responsible for managing the Institute's research program in education finance, governance, and human capital. He began working on education policy issues in the early 70s at the U.S. Office of Economic Opportunity, where his efforts were directed at the needs of poor and minority children.

Charles E. Metcalf is President of Mathematica Policy Research, Inc., which is one of the nation's leading independent research firms and conducts public policy research and surveys for federal and state governments as well as private clients. He is nationally known for his research on social experimentation and income distribution and has directed research activities at Mathematica for the past 21 years. Dr. Metcalf specializes in experimental and sample design, data collection design, and analytic design efforts. His expertise, gleaned from 28 years of experience in the field, spans all facets of research design and analysis. He has played a major role in more than 30 major social experiments, demonstrations, and evaluations. Dr. Metcalf has a Ph.D. in Economics from the Massachusetts Institute of Technology.

Morton Owen Schapiro is Professor of Economics and Dean of the College of Letters, Arts, and Sciences at the University of Southern California. He and Michael McPherson have co-authored two recent books on American higher education: *Keeping College Affordable: Government and Educational Opportunity* (Brookings 1991), and (with Gordon Winston), *Paying the Piper: Productivity, Incentives, and Financing in U.S. Higher Education* (University of Michigan Press 1993).

Fritz Scheuren has extensive experience in using administrative records in sample surveys and other settings. Currently, Visiting Professor of Statistics at The George Washington University, Dr.

Scheuren retired in 1994 as Director of the Statistics of the Income Division of the Internal Revenue Service. Formerly, he had been the Chief Mathematical Statistician at the Social Security Administration. In 1995, he won the Shiskin Award for contributions to U.S. economic statistics and among other honors is a Fellow of the American Statistical Association and the American Association for the Advancement of Science. He has published more than 90 papers, monographs, and books—both applied and theoretical—mainly in the area of survey sample design and estimation, including such topics as record linkage, privacy, and the handling of missing data. He holds a master's and doctoral degree in Statistics from The George Washington University.

Diane Stark is the Associate Director of the Center on National Education Policy. From 1988 to 1994, Ms. Stark was a legislative associate for the U.S. House of Representatives Committee on Education and Labor, where she assisted in the reauthorization of the major federal education programs. Prior to her work in the Congress, she was employed in the government relations offices of the National PTA and the Council of Chief State School Officers.

Cathleen Stasz is a Senior Behavioral Scientist at RAND and Site Director for the National Center for Research in Vocational Education (NCRVE). Her research areas include the implementation of advanced computer-based technologies in education, the workplace and the military, systemic school reform, and teaching and learning generic skills for the workplace. Currently, her projects include a study of the determinants of employer participation in school-to-work programs and an examination of the quality of student experiences in work-based learning environments.

Amy Rukea Stempel, Assistant Director for Standards Analysis at the Council for Basic Education (CBE), has been affiliated with the Council since 1989. In 1992, she left the CBE to teach the International Baccalaureate (English literature) at the Kodaikanal International School, Kodaikanal, India, and then returned to the CBE in the fall of 1994. Ms. Stempel has published numerous articles that inform the academic standards-setting process and the relationship of various education reforms to academic learning in CBE's flagship publication *Basic Education* and in *Teacher Magazine*. In addition, she designed the popular CBE chart "Standards: A Vision for Learning" (spring 1991), which synthesized all the current standards projects and was reprinted in 1994. A candidate for a master's degree in the Humanities at Georgetown University, Ms. Stempel is primarily engaged in writing about education reform and managing standards projects at CBE. She has a bachelor's degree in English from Carnegie Mellon University.

James W. Stigler is Professor of psychology at UCLA and Director of the Third International Mathematics and Science Study (TIMMSS) Videotape Classroom Study.

George Terhanian, a doctoral candidate at the University of Pennsylvania, is presently serving as an American Education Research Association Research Fellow at the National Center for Education Statistics. His general research interest lies in synthesizing evidence generated by local experiments

and nationally representative surveys. Mr Terhanian has several years of teaching and administrative experience in public and private schools.

DISCUSSANTS

Sharon Bobbitt received her Doctorate in Education Research from the University of Virginia in 1986. She worked for 8 years on the Schools and Staffing Survey, with a primary focus on teacher issues. She is currently Director of the Knowledge Applications Division in the Office of Educational Research and Improvement.

Dennis Carroll earned a Ph.D. in mathematics and quantitative psychology from Vanderbilt University in 1974. Since 1980, he has managed the National Center for Education Statistics' longitudinal studies program where several projects have incorporated records data.

Barbara S. Clements is Acting Director of the State Education Assessment Center at the Council of Chief State School Officers (CCSSO). She also directs the National Elementary/Secondary Education Data and Information System Project, which is funded by the U.S. Department of Education's National Center for Education Statistics to promote the standardization, automation, and effective utilization of data about education. Before joining the CCSSO staff, Dr. Clements worked on the development and administration of teacher assessment and evaluation instruments for the state education agency in Texas. She is a co-author of two textbooks on effective classroom management, soon to be released in their fourth edition. Dr. Clements holds a bachelor's degree in Education from the University of Texas at Austin, and is certified to teach secondary Spanish and Government. In addition, she has master's degree in Foreign Language Education and a Ph.D. in Educational Psychology from the University of Texas at Austin.

Emerson J. Elliott is a consultant on education policy, Federal statistics and management. He left the Federal Government in 1995 after a career that included heading the National Center for Education Statistics nearly eleven years and serving as the first "Commissioner of Education Statistics" when that post became a Presidentially appointed, Senate confirmed position under legislation enacted in 1988. Previously he had led the Issues Analysis Staff in the Office of the Under Secretary of Education, served as the Deputy Director of the National Institute of Education, and directed the OMB education branch when that was established in 1967.

Mary Frase is the Senior Technical Advisor in the Data Development and Longitudinal Studies Group, National Center for Education Statistics, in the U.S. Department of Education. Prior to joining NCES in 1985, she was a faculty member at Teachers College, Columbia University, and worked as an independent consultant advising state and local governments and conducting research in the areas of education policy, education finance, and state-local finance.

William H. Freund works within the U.S. Department of Education's National Center for Education Statistics. He recently assumed responsibility for adapting information technologies into the Center's data collections, program administration, and information dissemination. Just before this new position, he was responsible for institutional studies of postsecondary education. In that capacity, he was the program manager for the Integrated Postsecondary Education Data System (IPEDS)—a series of annual statistical surveys that collect enrollment, completions, finance, salary, and staffing data from the nation's postsecondary education institutions.

Paula R. Knepper is a Statistician in the Postsecondary Longitudinal Studies department of the National Center for Education Statistics.

James F. McKenney is currently the Director of Workforce Development, formerly the Office of College Employer Relations, at the American Association of Community Colleges (AACC). Also, he has served as the Assistant Vice President for Federal Relations, with responsibilities for the reauthorization of the Carl Perkins Vocational Education Act and the Job Training Partnership Act. As Director of Workforce Development, Dr. McKenney is charged with being the primary liaison between AACC and the various relevant federal departments and trade associations. In this role, Dr. McKenney has continued to track the implementation of the various federal human resource development laws. He received his bachelor's and master's degrees from the University of Florida and his doctorate from the University of Maryland.

Michael McPherson is the Dean of the faculty at Williams College. He is W. van Alan Clark Third Century Professor of Economics and Co-Director of the Williams Project on the Economics of Higher Education. Earlier, he served as Chair of the Williams Economics Department, as Senior Fellow in Economic Studies at the Brookings Institution and as Fellow of the Institute for Advanced Study. Mr. McPherson is co-author of two recent books, *Keeping College Affordable: Government and Educational Opportunity* (Brookings 1991) and *Paying the Piper: Productivity, Incentives and Financing in American Higher Education* (University of Michigan Press 1993). His new book, *Economic Analysis and Moral Philosophy*, co-authored with Daniel Hausman, was published by Cambridge University Press in 1996.

Jamie P. Merisotis is the founding President of the Institute for Higher Education Policy in Washington, D.C. The Institute is a non-profit, non-partisan organization with the mission of fostering access to and quality in postsecondary education through the development and promotion of innovative solutions to the important and complex issues facing higher education. The Institute has conducted a number of recent studies including *The Next Step: Student Aid for Student Success; College Debt and the American Family; Enhancing Quality in Higher Education;* and *Affirmative Action and the Distribution of Resources in U.S. Department of Education Programs.*

Kevin Miller is currently Associate Professor of Psychology at the Beckman Institute at the University of Illinois, Urbana-Champaign. His research interests concern the effects of symbolic

tools on cognitive development, focusing on how language and cultural differences between China and the United States affect the development of abilities such as reading and mathematical competence. He received his Ph.D. from the University of Minnesota, and then taught at Michigan State University and the University of Texas at Austin before joining the faculty at the University of Illinois. His research is currently supported by a Research Scientist Development Award and a research grant, both from the National Institute of Mental Health.

Frederick Mosteller is Roger I. Lee Professor of Mathematical Statistics Emeritus, Harvard University. He directs the Center for Evaluation of the Initiatives for Children Project at the American Academy of Arts and Sciences. Over the years, his research work has been devoted to theoretical and applied statistics. Dr. Mosteller works in data analysis, meta-analysis, robust methods, health and medicine, and social sciences, and has also written on sports statistics. While at Harvard, he has chaired the departments of Statistics, Biostatistics, and Health Policy and Management.

Mary Rollefson is a senior survey analyst with the National Center for Education Statistics. She has published several reports on teacher supply and demand and serves as the NCES liaison to the National Education Goals Panel.

Donald B. Rubin is Professor in the Department of Statistics, Harvard University. He has written nearly 250 publications (including several books) on a variety of topics, including computational methods, causal inference, survey methods, techniques for handling missing data, Bayesian methods, multiple imputation, matched sampling, and applications in many areas of social and biomedical science. Professor Rubin is a Fellow of the American Statistical Association, the Institute for Mathematical Statistics, the International Statistical Institute, the Woodrow Wilson Society, the John Simon Guggenheim Society, the New York Academy of Sciences, the American Association for the Advancement of Sciences, and the American Academy of Arts and Sciences. He is also the recipient of two of the most prestigious awards available to statisticians: the Samuel S. Wilks Medal of the American Statistical Association and the Parzen Prize for Statistical Innovation.

Eileen Mary Sclan is currently an Assistant Professor of Education in the Department of Curriculum and Instruction at Long Island University—C.W. Post Campus. Her main areas of research interest include teachers' workplace conditions, teacher performance evaluation, and teacher induction. At present, she is analyzing national data (funded by an AERA/NCES grant) to examine the inequitable distribution of qualified teachers and workplace supports. Dr. Sclan received her Ed.D. in Educational Leadership from Teachers College, Columbia University.

David Stern is Professor of Education at the University of California at Berkeley, and Director of the National Center for Research in Vocational Education, based at Berkeley's Graduate School of Education. From 1993 to 1995, he was principal administrator in the Center for Educational Research and Innovation at the Organization for Economic Cooperation and Development in Paris.

Since 1976, he has been on the faculty at Berkeley, teaching and conducting research on the relationship between education and work, and on resource allocation in schools. David Stern is the lead author of several recent books: *School to Work: Research Programs in the United States* (with N. Finkelstein, J. Stone III, J. Latting, and C. Dornsife 1995); *School-Based Enterprise: Productive Learning in American High Schools* (with J. Stone III, C. Hopkins, M. McMillion, and R. Crain 1994); and *Career Academies: Partnerships for Reconstructing American High Schools* (with M. Raby and C. Dayton 1992). He also co-edited *Market Failure in Training* (with J.M.M. Rtizen 1991), and *Adolescence and Work: Influences of Social Structure, Labor Markets, and Culture* (with D. Eichorn 1989).

P. Michael Timpane, Vice President of the Carnegie Foundation, is involved in developing all aspects of the programs of the Foundation. In his own research, he is assessing the progress and problems of contemporary national education reform. Mr. Timpane is also Professor of Education and former President of Teachers College, Columbia University, the world's most comprehensive graduate school for the preparation of educational, psychological, and health professionals. Previously, he served as Dean of Teachers College and as Deputy Director and Director of the federal government's National Institute of Education. He has conducted research on educational policy as a senior staff member at the Brookings Institution and the RAND Corporation. Also, Mr. Timpane is a member of the Pew Forum on Education Reform, for which he is currently organizing and editing a volume of essays on higher education's involvement in precollegiate school reform. In addition, he serves on the boards of Children's Television Workshop, the Southern Education Foundation, the Synergos Institute, and Jobs for Education and the American Associate of Higher Education. Mr. Timpane received a bachelor's and a master's degree in history from Catholic University, and an M.P.A. degree from Harvard University in 1970. He has received honorary doctorates from Wagner College and Catholic University.

Appendix B
Future NCES Data Collection Conference Agenda

THIS PAGE INTENTIONALLY LEFT BLANK

National Center for Education Statistics

Future NCES Data Collection: Some Possible Directions
Conference Agenda, November 27–29, 1995

Hyatt Regency Washington on Capitol Hill
400 New Jersey Avenue, N.W., Washington, DC 20001

Monday, November 27
7:00 p.m. Dinner in the Ticonderoga Room
Emerson Elliott will address the group after dinner.

Tuesday, November 28
8:30 a.m. Continental breakfast. Congressional B meeting room
Welcome and opening remarks by Jeanne Griffith.

9:00–10:30 a.m.
Session 1—Tracking Education Reform: Implications for Collecting National Data Through 2010
First Paper: Jack Jennings and Diane Stark
Second Paper: Chris Cross and Amy Stempel
External Discussant: Tom Kane
Internal Discussant: Mary Frase

10:30–10:45 a.m.—Break

10:45 a.m.–12:30 p.m.
Session 2—Curriculum, Pedagogy, and Professional Development
First Paper: Curriculum and Pedagogy: Implications for National Surveys
Authors: Cathy Stasz and Dominic Brewer
Second Paper: Teacher Education, Training, and Staff Development: Implications for National Surveys
Author: David Mandel
External Discussants: Michael Timpane and Eileen Sclan
Internal Discussants: Mary Rollefson and Sharon Bobbitt

12:30–1:30 p.m.—Lunch

1:30–2:45 p.m.
Session 3—Trends in Statistical and Analytic Methodology: Implications for National Surveys
Authors: Bob Boruch, George Terhanian, and Others

External Discussant: Fred Mosteller
Internal Discussant: Sue Ahmed

2:45–3:00 p.m.—Break

3:00–4:15 p.m.
Session 4—New Data Collection Methodologies, Part II: Experimental Design
Author: Chuck Metcalf
External Discussant: Don Rubin
Internal Discussants: Joe Conaty and Marilyn McMillen

4:15–4:30 p.m. Wrap-up first day

Wednesday, November 29—Congressional B meeting room

8:30–8:45 a.m.—Continental Breakfast

8:45–10:30 a.m.
Session 5—Postsecondary Education
First Paper: Tracking the Costs and Benefits of Postsecondary Education: Implications for National Surveys
Authors: Michael McPherson and Morty Schapiro
Second Paper: Special Issues in Postsecondary Education and Lifelong Learning: Implications for National Surveys
Authors: David Breneman and Fred Galloway
External Discussants: Jamie Merisotis and Jim McKenney
Internal Discussants: Roz Korb and Paula Knepper

10:30–10:45 a.m.—Break

10:45 a.m.–12:00 p.m.
Session 6—New Data Collection Methodologies, Part I: Observational Strategies
Author: Jim Stigler
External Discussant: Kevin Miller
Internal Discussant: Lois Peak

12:00–1:00 p.m.—Lunch

1:00–2:00 p.m.
Session 7—Education for Work: Curriculum, Performance, and Labor Market Outcomes
Author: Peter Cappelli
External Discussant: David Stern
Internal Discussants: Nabeel Alsalam and Marilyn Binkley

2:00–2:15 p.m.—Break

2:15–3:45 p.m.

Session 8—Using Administrative Records and New Developments in Technology

First Paper: Opportunities for Making More Effects Use of Administrative Records in Surveys of Elementary, Secondary, and Postsecondary Education

Author: Fritz Scheuren

Second Paper: New Developments in Technology: Implications for Collecting, Storing, Retrieving, and Disseminating National Data for Education

Author: Glynn Ligon

External Discussant: Barbara Clements

Internal Discussants: Dennis Carroll and Bill Freud

3:45–4:00 p.m.—Conference Wrap-up