

DOCUMENT RESUME

ED 370 894

SP 035 193

AUTHOR Singer, Judith D.; Willett, John B.  
 TITLE Methodological Issues in the Study of Teachers' Careers: Critical Features of a Truly Longitudinal Study.  
 PUB DATE Apr 94  
 NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 12-16, 1994). Contains small print.  
 PUB TYPE Speeches/Conference Papers (150) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120)  
 EDRS PRICE MF01/PC01 Plus Postage.  
 DESCRIPTORS \*Career Development; Data Collection; Elementary Secondary Education; Higher Education; \*Longitudinal Studies; \*Research Design; \*Research Methodology; Research Needs; \*Teachers; Teaching (Occupation)  
 IDENTIFIERS National Center for Education Statistics

ABSTRACT

This paper focuses on methodological issues in designing and conducting a longitudinal study of teachers' careers, as being explored by the National Center for Education Statistics (NCES). The paper outlines the need for data in four areas (teachers' worklives, teacher quality, the context of teaching, and teachers' career paths) and identifies six core design principles relevant to the study of teachers' careers. The six principles are: (1) collect truly longitudinal data; (2) view "time" as both an outcome and a predictor; (3) collect data on both time-varying and time-invariant measures; (4) collect data prospectively when necessary; (5) collect data in multiple base years; and (6) collect data from all relevant levels in the organizational hierarchy. The paper outlines several design alternatives that flow from these principles, evaluates the advantages and disadvantages of these alternatives, and prioritizes the concerns. The paper focuses on the topics of designating the target population and the specific sample to be tracked, as well as the length and periodicity of data collection. The paper suggests that NCES conduct several small-scale intensive data collection efforts--some of which could be embedded in the larger study, others of which should be conducted as pilot studies to be fielded before the full-fledged data collection effort. (Contains 15 references.) (JDD)

\*\*\*\*\*  
 \* Reproductions supplied by EDRS are the best that can be made \*  
 \* from the original document. \*  
 \*\*\*\*\*

*Methodological issues in the study of teachers' careers:  
Critical features of a truly longitudinal study*

*Judith D. Singer & John B. Willett  
Harvard University  
Graduate School of Education*

*Paper presented at the annual meeting of the  
American Educational Research Association  
New Orleans, LA  
April 1994  
12-16*

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

*J. D. Singer*

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

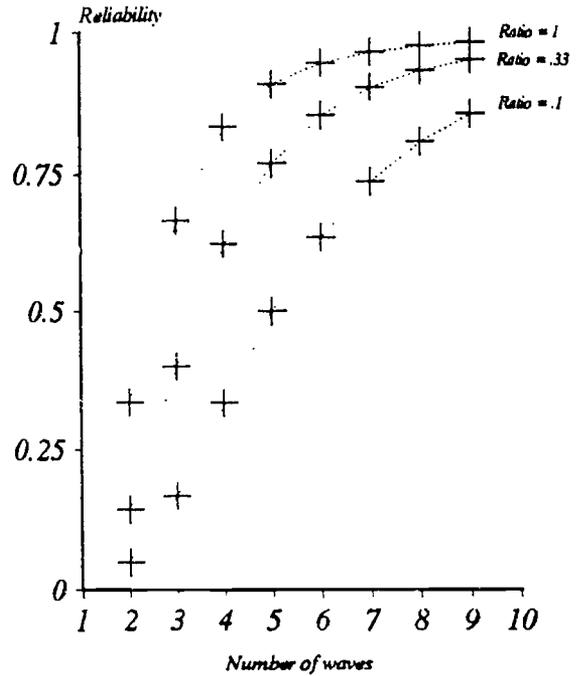
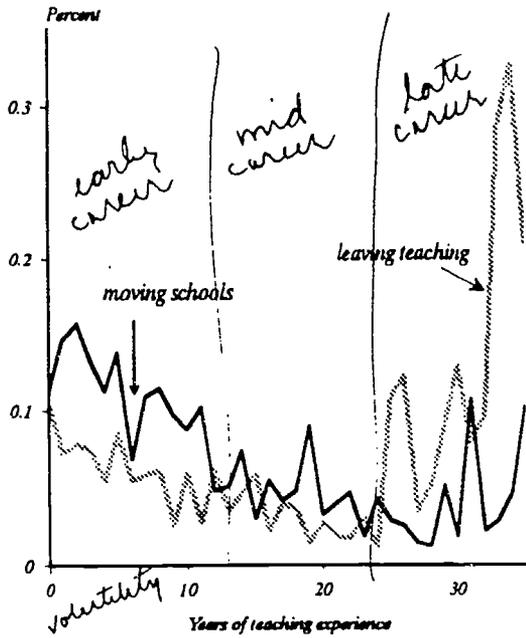
U. S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it
- Minor changes have been made to improve reproduction quality
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

*The order of the authors was determined by randomization. Please address comments to either author at Harvard University, Graduate School of Education, Larsen Hall, Cambridge, MA 02138, (617) 495-1961.*

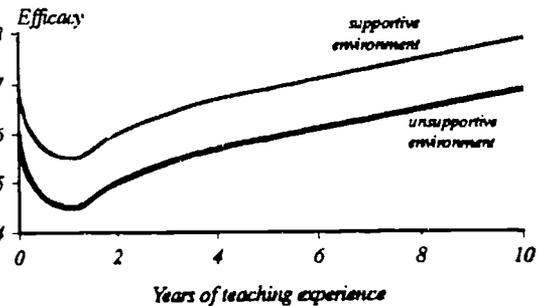
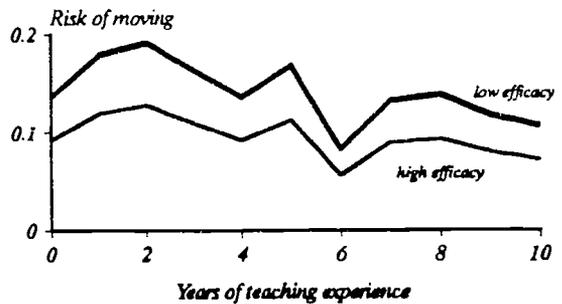
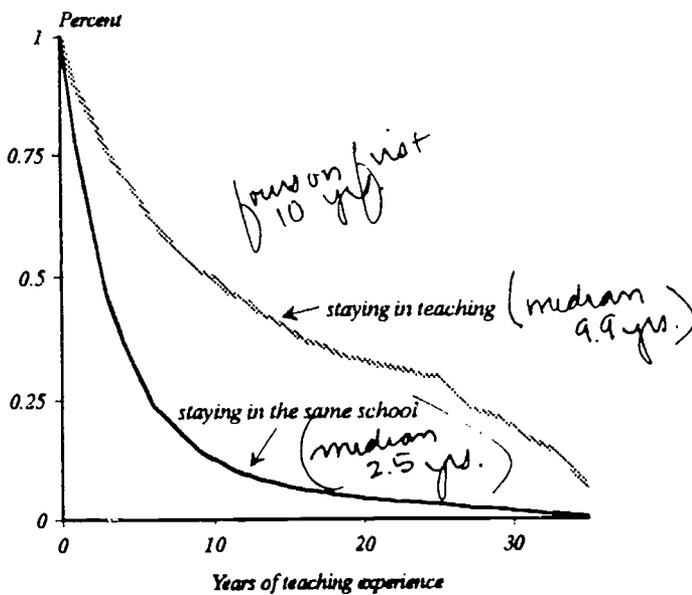
Reliability of change as a function of the number of waves of data collection

Estimated risk of leaving teaching or moving schools



Source: Willett (1988) Educational and psychological measurement, Vol 49, 587-602.

Estimated survivor functions for staying in teaching or staying in the same school



Source: Singer & Willett (1991) Psychological Bulletin, Vol 110, 268-290.

Based on: Chester (1991) Unpublished doctoral dissertation, HGSE.

a 1/4 th or a 1/3 rd of  
those who k remain

## Methodological issues in the study of teachers' careers: Critical features of a truly longitudinal study

Judith D. Singer & John B. Willett<sup>1</sup>  
Harvard University  
Graduate School of Education

The National Center for Education Statistics (NCES) is exploring the possibility of conducting a large-scale multi-year study of teachers' careers. Unlike their two current survey programs for teachers--the cross-sectional Schools and Staffing Survey (SASS) and its companion one-year prospective Teacher Followup Survey (TFS)--this new study would follow a national probability sample of teachers over an extended period of time. Recognizing that this effort would involve considerable expense and time, NCES has taken several steps towards evaluating the need for, and the feasibility of, such a data collection enterprise.

As part of this process, NCES commissioned a panel of experts to write papers presenting their perspectives on the *substantive* issues involved in a longitudinal study of teachers' careers. Taken together, these four papers (Billingsley, 1992; Grissmer & Kirby, 1992; Theobald & Gritz, 1993; and Weiss, 1992) substantiate the need for a longitudinal study and argue effectively why NCES should consider undertaking this major project during the late 1990s. At a January 1993 planning conference in Washington, DC attended by these experts and other researchers and policymakers from both within and outside of NCES, however, it became clear that there were a number of *methodological* issues that had to be addressed before initiating the study. Precisely which teachers should be studied? Should NCES follow a probability sample of all teachers at varying points in their careers or should they focus on teachers at a particular career juncture, say beginning teachers, mid-career teachers, or perhaps teachers nearing retirement? For how long should the teachers be followed? Is complete follow-up until retirement necessary or would a shorter time-period suffice? How often should teachers be contacted? Is it necessary to collect data every year, or is a shorter interval necessary or a longer interval sufficient?

After this conference, NCES asked us to write a paper focusing on these and other methodological

issues that arise in the consideration of how to design and conduct a longitudinal study of teachers' careers. This paper, the result of that request, builds on the results of the January 1993 planning conference, our subsequent discussions with members of the NCES staff, and our previous work on teachers' careers and research design. In it, we describe the purposes of this longitudinal study, identify six core design principles relevant to the study of teachers' careers, outline several design alternatives that flow from these principles, evaluate the advantages and disadvantages of these alternatives, and prioritize the concerns. We focus most closely on the topics of designating the target population and the specific sample to be tracked, as well as the length and periodicity of data collection. In addition, we raise a number of measurement issues that will arise as NCES considers expanding its data collection efforts for teachers beyond their current cross-sectional and one-year follow-up studies. In the process, we suggest that NCES conduct several small-scale intensive data collection efforts--some of which could be embedded in this larger study, others of which should be conducted as pilot studies to be fielded before the full-fledged data collection effort.

In identifying what we see as the major methodological concerns for the design of a longitudinal study of teachers' careers, we have purposefully omitted the discussion of several common design issues, except insofar as they arise during the discussion of one of our focal topics. In particular, we do not discuss issues of statistical power and sample size, nor do we discuss the practical issues involved in listing the target population and actually drawing the desired probability sample. It is not that we consider these topics unimportant; rather, we believe that NCES has the in-house talent who can best address these practical issues *after* dealing with what we consider the larger conceptual issues outlined in this paper.

---

<sup>1</sup> Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 1994. The order of the authors was determined by randomization. We wish to thank Dan Kasprzyk and Sharon Bobbitt of the National Center for Education Statistics for their valuable insights concerning this project. Please address correspondence to either author at Harvard University, Graduate School of Education, Appian Way, Cambridge, MA 02138.

### Research context and goals:

#### Why should NCES conduct a longitudinal study of teachers' careers?

During the past decade, as the quantity of quantitative research on teachers' careers has grown, the quality has improved as well. The empirical literature, once comprised solely of annual attrition rates computed at either the district or state levels (see, e.g., Darling-Hammond, 1984) now includes large-scale in-depth studies of who stays in teaching, who leaves teaching, who returns to teaching, and when and why these transitions occur (Willett & Singer, 1991). As we contemplated the initiation of a future longitudinal study, we considered the quality of the research base that already exists, what the research community has learned from this literature, the knowledge gaps that remain unfilled, and the contribution that an NCES data collection effort would be able to make.

#### What researchers currently know about teachers' careers

Most large-scale quantitative research on teachers' careers has relied on data collected from one of four sources:

- *state-level administrative records*, which are used to reconstruct longitudinal career histories of entire cohorts of newly-hired teachers in a single geographic area (see, e.g., Grissmer & Kirby, 1992; Murnane, Singer, & Willett, 1988, 1989; Schlechty & Vance, 1981; Theobald, 1990);
- *national probability samples of college graduates*, which are used to construct longitudinal career histories of individuals who ever taught anywhere in the United States (see, e.g., Hafner & Owings, 1991; Heyns, 1988; Murnane, Singer, Willett, Kemple, & Olsen, 1991);
- *retrospective reports from national probability samples of current teachers* who participate in NCES' tri-annual SASS, which are used to describe teachers' previous career decisions and their future intentions (see, e.g., Bobbitt, Faupel, & Burns, 1991); and,
- *one-year followup reports from national probability subsamples of teachers* who participate in the NCES' tri-annual TFS, which are used to estimate annual attrition rates, to identify the career decisions of teachers who participated in SASS, and to describe teachers' current work conditions (see, e.g., Choy, Medrich, Henke, & Bobbitt, 1992).

Although there have also been many smaller scale intensive studies of teachers in individual school districts, it has been the increasing number of studies in these four traditions that have provided the most generalizable evidence concerning teachers' careers.

What have we learned from these bodies of literature? Although a review of this vast literature is

beyond the scope of the present paper, several key findings have emerged (for a recent review, see the National Academy of Sciences 1992 report *Teacher Supply, Demand, and Quality: Policy Issues, Models and Data Bases*). These data bases have allowed researchers to document that the appeal of teaching declined precipitously during the 1970s and 1980s, that the first years in teaching continue to be the riskiest for all teachers, and that the schools still lose those individuals who score particularly well on standardized tests, who have attractive career opportunities outside the schools, or are paid comparably lower salaries. A further lesson has been the great mobility of the US teaching force: not only do many teachers who leave teaching ultimately return, the reserve pool of former teachers—not the pool of recent college graduates—now comprises the major source of teacher supply.

#### Prominent knowledge gaps in the study of teachers' careers

Despite the fact that researchers have matured beyond the calculation of simple attrition rates, developing a deeper understanding of issues underlying the relationship between teacher supply and demand, the knowledge base for describing *teachers' worklives* remains largely descriptive. Few large-scale studies have delved into the reasons *why* teachers behave the way they do. Many questions remain unanswered. Building upon the results of the January 1993 conference and the 1992 report of the National Academy of Sciences, below we identify what we believe are the four most prominent areas in which additional data are needed.

**Teachers' worklives.** One fundamental knowledge gap in the study of the US teaching force concerns an understanding of teachers as workers, leaders, mentors, and individuals and how these roles evolve and change over time. The currently available longitudinal data sources (primarily administrative records and large multi-purpose national surveys) have not included the kinds of information researchers need to describe fully teachers' lives in schools. And although NCES' two current data collection efforts on teachers provide some insight into these topics, these studies are not truly longitudinal, thereby describing status, not change. The clear consensus from the January planning conference was that a major contribution of a longitudinal study would be the collection of detailed data describing teachers lives in school. In particular, the study should gather data about teachers' work roles and working conditions and how these features vary across settings and change over time (Billingsley, 1992). Under this broad rubric falls the study of topics such as the on-going support and development of teachers as professionals and leaders; teachers' perceptions of their work climate,

reform initiatives, school administrators, and operations; and teachers' job commitment and job satisfaction. Although the current SASS and TFS attempt to gather measures of some of these constructs, the questions are brief, primarily because the studies currently focus on issues such as degree attainment and certification. By gathering longitudinal data on teachers' worklives, researchers might be able to determine *which* teachers are most likely to leave, and whether certain characteristics of jobs and schools are associated with lower attrition, higher satisfaction, and stronger commitment.

**Teacher quality.** A second knowledge gap—and an inevitably difficult one to fill—is that no study of teachers' careers has yet attempted to measure the "quality" of the nation's teaching force. Issues of teacher quality are thorny (to say the least) and a discussion of the adequate measurement of this construct is beyond the scope of this paper (Kennedy, 1992). But what is relevant for consideration at this time is the simple fact that a new longitudinal data collection effort focused on teachers should attempt to gather some data on teacher quality. It is only through the measurement of this elusive construct that we will begin to know whether the teachers who leave the schools are perhaps the very ones the schools might want to leave—those of lower quality.

In this paper, we will not wade into the firestorm of substantive, political, and methodological controversy surrounding the measurement of teacher quality. We will address the topic only through its ramifications for other aspects of design, as when we discuss the need for data collected from individuals other than the teachers themselves. Because of the importance of this topic, however, we recommend that NCES contact experts in the field and work with them to explore possible strategies for collecting data on teacher quality. The state of knowledge about teachers' careers is such that the issue of "quality" is increasingly emerging as a fundamental question (Shulman, 1992). To conduct a truly longitudinal study of teachers' careers without attempting to measure some aspects of quality would be, to our mind, a grave omission.

**Teachers' work contexts.** The current SASS and TFS gather only limited data on teachers' work contexts. Administrators, principals, and colleagues respond to questionnaires, but they are asked general questions about their schools, not specific questions about the particular teachers who were also surveyed in the study. Teachers, too, are asked to describe their perceptions of administrators, colleagues, and students, but with the exception of the administrator questionnaire, direct linkage is difficult, if not impossible. A longitudinal study would provide an ideal setting in which to delve further into this topic. Who are the administrators with whom teachers work?

How do these administrators view these teachers? And who are the students the teachers serve? Such linkage might allow us to expand greatly our knowledge concerning *who* is really leaving teaching and who returns to teaching after a career interruption.

Under the rubric of teachers' work contexts, we raise the possibility of a further data source: the students the teachers serve. A truly longitudinal SASS would be an ideal setting in which to gather data on the students in the classes of the surveyed teachers. Although data on teachers has been gathered in other NCES data collection efforts (e.g., NLS-72, HSB, NELS) the focus of these studies has been the students, not the teachers. The students have remained constant across waves while the teachers changed. In a longitudinal study of teachers, it would be possible to turn the tables and have the tracking take place in the opposite direction—having the teachers remain constant across waves while the students change. This would allow researchers to determine, for example, whether changes in teachers' commitment are related to changes in the students they serve. Do teachers leave teaching when they have to teach more difficult students?

**Teachers' career paths.** Fourth, despite the fact that there has been more research on this aspect of teachers' careers than on any other, substantial knowledge gaps persist. Although most researchers recognize that teachers' careers unfold over time, most studies still rely on data collected retrospectively, cross-sectionally, or at only two points in time. No study has yet to juxtapose career decisions alongside full data on wage and benefit histories as well as workplace conditions and family demands. No study has attempted to track the labor market experiences of current teachers, former teachers, and returning teachers. No study has attempted to follow newly licensed teachers as they enter teaching, leave teaching, return to teaching, and perhaps leave once again. As we argue elsewhere (Willett & Singer, 1991), it is only through the tracking of teachers through their various career paths that we will be able to truly understand when and why teachers make the many career transitions that they do.

The need for better data in these four areas: teachers' worklives, teacher quality, the context of teaching, and teachers' career paths—are at the core of our recommendations. These are not the only four areas that should be addressed in a longitudinal study of teachers' careers, but we have highlighted them because they represent, to our view, the four most important focuses for a future longitudinal study and because taken together, they have direct implications for research design. When we note, for example, that NCES should collect data on the context of teaching, this implies that data be collected from the people with whom teachers work. Because of the direct link between research goals and research design, we now

discuss the design implications of these four substantive goals.

#### Implications of the research goals:

##### Six core design principles

We began this project by wanting to specify, in great detail, the specific components of a longitudinal study. We hoped to be able to recommend that NCES follow  $n$  teachers for  $y$  years, collecting data every  $m$  months. This had a satisfying precision, and at first glance, would seem to help NCES most directly by giving specific recommendations to which staff could react.

But when we attempted to formulate such recommendations, we discovered that most design decisions were inextricably linked. There was no single optimal design, best suited to all purposes. Decisions had to be linked to discussions about research goals. Compromises had to be struck. Rather than offering a single design, we therefore decided that a more helpful strategy would be to: (1) identify a set of core design principles that flow from the research goals; (2) specify the issues involved in considering various approaches to implementing each of these principles; and (3) offer recommendations based on consideration of the advantages and disadvantages of each approach. We begin this process in this section, by presenting six principles that we believe flow from the substantive data needs identified above. In subsequent sections, we will describe the ramifications of these principles. By beginning with general principles instead of specific recommendations, we hope that if NCES staff decide that, for reasons of cost or scope, they cannot adhere to our specific recommendations, they can still consider these general principles as they shape a design of a different structure.

##### Design principle 1: Collect truly longitudinal data

Too little of the research on teachers' careers is based upon studies that have followed the careers of teachers. Instead, much of what we know about teachers' careers has been gathered from studies based on cross-sectional designs. Without truly longitudinal data, we have no way of studying change (Rogosa, Brandt, & Zimowski, 1982; Willett, 1988). If an analysis of the SASS reveals, for example, that teachers with more years of experience have lower levels of commitment than teachers with fewer years of experience, we cannot infer that teachers' commitment decreases with experience. The teachers with more years of experience differ in important and fundamental ways from the teachers with fewer years of experience—they graduated from college in different years, they were licensed in different years, they entered teaching in different years, and they have taught under types of working conditions.

Information about status in one year may not tell us anything about how the teachers got to this particular point. Perhaps the commitment of many mid-career teachers has actually *increased* over time, and that the low level we see in a cross-sectional survey is *higher* than we would have seen had we collected data on these teachers during their previous years on the job.

Recognizing the limitations of cross-sectional data, the current NCES data collection program includes the Teacher Follow-Up Survey, conducted one year after each base year SASS. Although the TFS is helpful for computing attrition rates and for studying the short-term mobility of the teaching force, two waves of data are inadequate for studying change.<sup>2</sup> Two waves of data separated by only one school year cannot portray the complex patterns of growth and change we expect teachers to display. Among the many problems with two-wave designs is their omission of data required for characterizing the *shape* of each teacher's growth trajectory. Two waves of data tell researchers only about each teacher's status at two points in time; there is no information about *how* the teachers got from point A to point B. In that single year, did all the change occur immediately after the beginning of the school year? Did most teachers change at a steady pace, at equal amounts each month? Did some teachers remain at a steady pace for most of the year, only to fall off in commitment by the spring? What will happen as these teachers continue to teach? With two waves of data separated by only one school year, it is impossible to know.

We recommend that NCES expand their longitudinal data collection in two ways: (1) by collecting more waves of data; and, (2) by extending the duration of data collection over a longer period of time. In a later section of the paper, we offer specific recommendations concerning these two principles by asking: How many waves should be collected? Over what period of time? Spaced at what intervals? For now, however, let us comment broadly on the impetus behind these two recommendations.

**Collect as many waves of data as possible.** It is well known that the best way to comment cogently on change over time is to collect data at three or more timepoints. With three or more waves of data, a researcher can construct a more fine-grained photograph of each teacher's development over time. Patterns of change can be seen even in simple plots of status versus time. Is change linear or curvilinear? Do most teachers peak after their first few years on the job, or do many increase steadily over time? Are increases

<sup>2</sup>We use the word change here in its broadest sense, to encompass changes in such diverse domains as educational status, family status, employment status, and attitudes, knowledge, and behavior.

in one dimension (say, self-perceived competency) accompanied by decreases in other dimensions (say, job stress)? Are teachers with declining levels of efficacy more likely to leave than those on an upward trajectory? Do most changes occur during the early years on the job and the years near retirement, or are there substantial change during mid-career as well?

Answers to such questions about changes in teachers' worklives, quality, work contexts, and career paths require more waves of data than NCES currently collects. Three waves are the minimal set of data for making inferences about change; four or five waves would be far better. Within a set of fiscal and logistical constraints, of course, advice such as "the more, the better" is unrealistic. Issues including the total duration of data collection and the periodicity of data collection also must enter into consideration. At a global level, however, our recommendation is clear: **NCES must commit to following samples of teachers over at least three waves of data collection.**

**Maximize the length of the data collection period.** The second consequence of Design Principle 1 is that the duration of this longitudinal study must be long enough to permit the careful study of change over time. Teachers are not students; the changes they will exhibit are likely to be more subtle and slow. Attrition rates are low. Yes, teachers move between schools and school districts, there are pockets of great turnover and change, and there are increasing percentages of teachers nearing retirement, but as a whole, the nation's teaching force is relatively stable in comparison to years past. The study of change in a relatively stable environment requires long longitudinal data sequences. In addition, it is only with such long records that will researchers be able to address important policy questions concerning returning teachers. The reason is simple: For a longitudinal study to describe returning teachers, data must be collected over a long enough period of time so that the teachers who leave will have time to return.

If there were no need for immediate results, we might recommend that NCES collect data over the *entire* teaching career—from major choice in college through retirement from the profession. "Cradle to grave" data collection is ideal for studying change over time. It would allow us to learn, for example, whether career patterns and levels of commitment differ for teachers who appeared committed to an education career in their college years in comparison to other teachers who came to the profession at a later point in time. Or whether teachers we might have been able to identify, on the basis of early signs and signals, were likely to leave teaching, never to return.

"Cradle-to-grave" data collection has another appeal as well: it ensures that researchers capture *change whenever it occurs*. The current state of knowledge about teachers' careers provides the

research community with only the most limited information about when and why changes occur. If NCES were able to collect data throughout the entire teaching career, researchers would be sure not to miss any of the crucial times of change or transition. As importantly, the longitudinal data record preceding the transition could be used to model whether and when transitions occur.

"Cradle-to-grave" data collection is, of course, impossible. Thus, NCES must consider whether they would be best served by one of several more realistic options:

- **following all teachers teaching in a given year over time.** This would allow NCES to follow all teachers who participated in a given base year of the SASS for an extended period of time. Generalizability would appear to be great. However, the ability to make generalizable inferences would be limited by the initial use of a "stock sample" (Lancaster, 1990). By selecting an initial sample of teachers who differ widely in their initial career stages, it would be difficult to reconstruct full career histories for everyone under study. Moreover, it would be virtually impossible to separate out the effects of age, period, and cohort (an issue discussed in a later section). We put forth this option not because we recommend it, but because it was the assumed plan offered by NCES staff at the January 1992 conference.
- **following selected cohorts of teachers with different amounts of experience over time.** This would allow NCES to sample current teachers with say, 1, 6, 11, 16, 21, and 16 years of experience and follow them for an extended period of time. The vintages of each of the initial cohorts would be staggered (in accordance with the periodicity of data collection and length of data collection) to permit overlapping "age" periods. This would allow researchers to obtain information about most years of the career from more than one cohort, thereby helping to unravel the age-period-cohort problem. For example, data describing the 6<sup>th</sup> year of the teaching career would be obtained from two distinct samples: in a later data collection wave for the 1<sup>st</sup> year cohort and in an initial data collection wave for the 6<sup>th</sup> year cohort.
- **selecting a single period of the teaching career for intensive study and following a single cohort of teachers over time.** NCES would sample teachers in a crucial transition phase, perhaps the first 5 years in teaching and follow this more select cohort over an extended period of time. NCES would need to decide which phase of the teaching career—early, mid, or late—deserved this intensive focus. It would also need to make some decisions concerning returning teachers—would a teacher in

the first year of a second spell be considered a "beginning teacher." This approach would produce the clearest results, in the sense of homogeneity of the initial sample and special clarity if the group chosen was beginning teachers, but generalizability would be far more limited. A second problem would arise if the age cohort chosen happened to experience great stability, not transition. In this event, there would be little change to study.

Each of these approaches—representative of three broad classes of longitudinal design—has advantages and disadvantages, which we will discuss in a later section. At a global level, however, whichever approach is adopted, our recommendation remains clear: **NCES must commit to following the sampled teachers for a long enough period of time so that researchers will observe the changes of interest.**

**Design principle 2: View "time" as both an outcome and a predictor**

Most longitudinal studies of teachers' careers (of most individuals, in fact) have viewed chronological time as either an outcome or as a predictor.<sup>3</sup> This bifurcation has arisen as an artifact of both disciplinary boundaries and substantive focus. Studies examining whether and, if so, when teachers leave teaching—arising primarily from a tradition of economics—have seen time as the outcome variable, an object of study in its own right. Researchers examine *whether* the teachers experience particular transitions (entering teaching, moving to another school or school district, leaving teaching, or returning to teaching) and *when* these transitions occur. Studies examining changes in teachers' attitudes, knowledge, or behavior—arising from disciplines such as sociology and organizational psychology—in contrast, have treated time as a predictor. Researchers in this tradition study *whether* and *how* attributes of teachers change over time.

We believe that both perspectives deserve equal voice when designing a longitudinal study of teachers' careers. It would be an error, we believe, for NCES to prioritize these perspectives, placing *only* the study of the movement in and out of teaching (time as the outcome) in priority over the study of changes over time (time as a predictor). In the past decade, there has been a great deal of policy interest in issues of teacher supply and demand. Hence, studies that have treated time as the outcome have received more attention than have studies that have treated time as a predictor. But at the January 1993 planning conference, Emerson Elliot, Commissioner of NCES

made it very clear that in the years ahead, increasing attention needed to be given to the study of changes in teachers' worklives over time.

Decisions about number of waves, spacing of waves, and length of data collection are usually made within the context of only one of these perspectives. Researchers interested in measuring growth, for example, treat time as a predictor and make design recommendations based on this view (see, e.g., Willett, 1988). The goal of design, from this perspective, is to determine the length of data collection and the spacing of waves so that a statistical summary of *growth* can be computed as precisely as possible. Researchers interested in measuring event occurrence, in contrast, treat time as an outcome and make design recommendations from that point of view (see, e.g., Singer & Willett, 1991). The goal of design, from this perspective, is to determine the length of data collection and the spacing of waves so that a statistical summary of *event occurrence* can be computed as precisely as possible. As described in a later section of this paper, on occasion, both perspectives lead to the same design recommendation; on other occasions, however, they conflict.

It is our view that **the design of a multi-purpose longitudinal study of teachers' careers must not favor one perspective to the exclusion of the other.** The practical implication of this view is that NCES may have to commit to collecting data more frequently and for a longer period of time than they might have chosen had they decided on designing this longitudinal study from only one perspective. If, for example, the "time as an outcome" perspective would allow a bi-annual data collection schedule, but the "time as a predictor" perspective points to a *semi-annual* schedule, we would recommend the more frequent periodicity. Our reason is simple. Use of the more frequent periodicity would certainly not harm researchers working in the first tradition, but failing to do so would certainly stymie researchers operating in the second tradition.

Further reflection on this topic also suggests that researchers from the two traditions may not be nearly as at odds with each other as one might initially suspect. We began our discussions of design by considering these two perspectives as separate and distinct. Even a simple review of the literature reinforces this stereotype. Yet we now see the possibility of a merger between traditions because of a simple insight: **one tradition's outcome is the other tradition's predictor.** Those attributes in which researchers are interested in measuring change over time (teacher self-esteem, relationship with principals) are those attributes that these same researcher would certainly consider using as time-varying covariates for the study of the teacher career, if they had them available. The problem historically, however, has been

<sup>3</sup>See, for example, the special issue of the *International Journal of Educational Research* focused on teachers' careers edited by Huberman, 1988.

that studies in this tradition have not collected data from the teachers themselves, precluding the investigation of such effects. Conversely, the events whose occurrence researchers are studying (transferring from one school to another) may affect a teacher's growth trajectory, producing important and measurable impacts on the level, shape or curvature. But here, too, the unavailability of longitudinal records on event occurrence has precluded the investigation of these sorts of effects.

Our review of the literature suggests that lack of relevant data has allowed these two research traditions to grow in isolation. A major contribution of a truly longitudinal study of teachers' careers would be the potential synergy that would come from researchers working in both these traditions using the same data resource.

#### **Design principle 3: Collect data on both time-varying and time-invariant measures**

All the data that will be collected in this longitudinal study can be classified as either time-invariant or time-varying measures. As their name implies, time-invariant measures are characteristics of individuals that do not change over time. In the study of teachers careers, information on demographic characteristics such as year of birth, sex, year of college graduation, year of licensure, and college major do not change over time. Data on time-invariant measures need to be collected only *once* during a longitudinal study. By their very nature, the values of time-invariant predictors do not change. If such information is collected during the base year of the longitudinal study, there is no need to collect these data again. This frees precious data collection time and allows it to be devoted to the acquisition of time-varying information.

Time-varying measures are also true to their name: they are variables whose *values* vary over time. Teacher efficacy, class size, salary, working conditions, and family composition are some of the many time-varying measures that this longitudinal study will need to collect. Data on time-varying predictors are best collected over time. Although it is possible to retrospectively construct the values of some time-varying measures (see the next design principle for a further discussion), data will be gathered with greater precision and validity if the values of time-varying measures are gathered *as they vary*. Can a researcher reasonably expect a teacher with ten years of teaching experience to reliably and validly retrospect back to his or her first year in the classroom? If researchers need to study the variation in time-varying measures (in studies of growth) or if they want to use time-varying measures as predictors in our studies of whether and when teachers transfer from a school or leave teaching, NCES must collect data on these

variables as they unfold over time.

This measurement issue has a direct implication for the spacing of data collection waves. If the values of time-varying measures change often, data collection waves must be spaced closely together. If they are spaced too far apart, teachers will need to retrospect far back to construct their answers. If the values of time-varying measures change infrequently, however, data collection waves may be spaced further apart. After all, if answer change infrequently, the measure becomes essentially "time-invariant" during the interview periods.

Of course, in the design of a large-multi-purpose study, there will be some time-varying predictors that change frequently (teacher efficacy, for example, may change on a weekly or even moment-by-moment basis), while others will change relatively infrequently (class size or types of students served may change only on a semester or annual basis). This may suggest that NCES consider using different data collection periodicities for different measures. It may be cost-effective and efficient, for example, to collect some types of data on a semi-annual basis, and other types of data on an annual or bi-annual basis. In a later section of this paper, we describe the implications of the rate of change in time-varying predictors for the spacing of waves. For now, however, we note the global implication of this issue: **NCES must commit adequate resources to measuring the values of time-varying variables as often as necessary.**

#### **Design principle 4: Collect data prospectively when necessary**

From one perspective, most of the data that NCES collects on its teacher questionnaires are retrospective. When a teachers is asked to rate whether student violence is a problem in his or her school, the teacher is reflecting (retrospecting) on his or her experience. A researcher can attempt to limit the time frame involved in the retrospection by adding a time-frame to the question. In some instances, the item might be worded to ask "*During the past week,...*" whereas in others, the item might include the time frame "*During this school year,...*"

Retrospective data collection is fraught with problems. First consider the easiest types of data to collect retrospectively—event occurrence. Although rare events—college graduation, first teaching job, entering this school district—may be remembered indefinitely and highly salient events—a leave of absence from teaching—may be remembered for several years, habitual events—daily work activities—are forgotten almost immediately (Bradburn, 1983; Sudman & Bradburn, 1982). The longer the time period, the greater the error. Three types of errors are especially common: (a) *memory failures*, in which respondents forget events entirely; (b) *telescoping*, in

which events are remembered as having occurred more recently than they actually did; and (c) *rounding*, in which respondents drop fractions and report even numbers or numbers ending in 0 and 5. These errors create different biases: Memory failures lead to underreporting, telescoping to overreporting, and rounding to both.

If gathering retrospective information about event occurrence is difficult, gathering retrospective data that require qualitative or quantitative judgments is far more difficult. When you ask teachers to retrospect about "states"—attitudes towards their jobs, efficacy, satisfaction, or commitment—the errors escalate. It is virtually impossible to collect retrospectively reliable and valid attitudinal and affective data (cite).

The implication of this recognition is clear: NCES **must commit itself to collecting prospective data whenever necessary to ensure the validity and reliability of responses.** As we discuss in a later section, adherence to this principle may require increased periodicity of data collection for certain types of information. Data may be collected retrospectively only when it will yield as reliable and valid data as if it were collected prospectively.

There is, however, one area of retrospective data collection for which we recommend that NCES devote considerable time and energy: the retrospective reconstruction of prior work and education histories among all teachers in the base year sample. Regardless of how NCES ultimately identifies the sample of teachers to be included in this longitudinal study (by using everyone in the base year SASS or just a sample of teachers at different points in their careers), each teacher in the study will have already experienced a variety of events that will be important to consider when modeling subsequent career behavior. Elsewhere (Willett & Singer, 1991), we have recommended that the study of teachers' careers begin, whenever possible, at the beginning of teachers' careers, but it is clear that for the present study, this ideal will not be attained. As a result, each sampled teacher will have "a story"—an important part of his or her data collection record. It is imperative that NCES use strategies during the initial wave of data collection that allow for the rigorous construction of these career histories. The present set of questions included in the SASS are totally inadequate for constructing these histories. It would be a grave error, in our view, to collect prospective data without first placing each teacher on his or her own career timetable.

#### **Design principle 5: Collect data in multiple base years**

In its original plan for SASS, NCES wisely recognized that the teaching force changes over time. The decision to field a national survey every three years (with a smaller follow-up after each round of

data collection) ensures that time-related differences in the teaching force can be registered. By comparing the percent of fully certified teachers in 1987 with the percent of fully certified teachers in 1990 and 1994, it would seem that researchers can comment cogently on differences between the teaching force in one year and the teaching force in another. This issue is of great interest to policymakers concerned with the quality of our nation's teaching force.

We recommend that NCES make a similar resource commitment in the design of this proposed longitudinal study by having the initial years of data collection be staggered across multiple base years. In other words, we recommend that NCES *not* conduct this longitudinal study by following a cohort of teachers teaching in only one base year, but rather that NCES use several base years for identifying the initial group of teachers to follow. Our recommendation is similar to the strategy that NCES employed when their data collection program for secondary school students moved from the single cohort NLS-72 to the dual cohort HSB.

Recognizing the cost implications of this recommendation, we do not make it lightly. We believe strongly, however, that if NCES is to conduct this longitudinal study, they would be best served by doing it well, recognizing the rival hypotheses that will plague data analysts in the years ahead. The inclusion of multiple base years is, in our opinion, the best way that NCES will prevent researchers from being totally stymied by a technical difficulty known by methodologists as the "Age-Period-Cohort" problem (Fienberg & Mason, 1987; Baltes, Nesselroade, Shaie). Below we give a brief overview of the problem; in a later section of the paper, we delve into its implications in further detail.

When a researcher describes a teacher's place in his or her career, the teacher's "place in time" can be measured using each of the following three markers: by his or her entry cohort (the year the teacher started teaching), by the year of his or her career (1<sup>st</sup> year, 2<sup>nd</sup> year, known more generally as "age"), and by the chronological year when data being described (1995, 1996, known more generally as "period"). All three time metrics can be substantively important. Teachers may have certain attitudes or go through certain transitions because of any one of them. Teachers may be affected by the particular year they entered teaching (the cohort effect), the particular year of their career (the age effect), and by the particular year that is being described (the period effect).

Setting aside the obvious complications that arise for teachers who have had breaks in service, **knowledge of any two of these time markers fully defines the third.** If, for example, a researcher collects data on the 3<sup>rd</sup> year of the career for a teacher who entered in 1990, the chronological year being described

must be 1993. Or if the chronological year of data collection is 1996, then data describing 1<sup>st</sup> year teachers will only describe those in the 1996 entry cohort and data describing 2<sup>nd</sup> year teachers will only describe those in the 1995 entry cohort. This confound makes it difficult to determine whether teachers behave the way they do because of their entry cohort, year of the career, or the particular time period being referenced.

Research on teachers' career paths have been plagued by this problem. When Mark and Anderson (1985) published one of the first longitudinal studies of teachers' careers, they concluded that teachers hired in the mid-to-late 1970s were much less likely to stay in teaching than teachers hired in the same school districts before them. But in a reanalysis of these same data, Singer and Willett (1989) showed that this supposed "entry effect" was more likely attributable to a "period" effect—teachers hired in the later cohorts were subjected to RIFs in the early 1980s, thus giving them the appearance of having "less commitment" but actually their departure appeared to be more a function of something else.

Cross-sectional data collection totally confounds all three sources of information about time. When a researcher analyzes the 1990 SASS and finds differences across teachers with different years of experience, it is impossible to know whether these differences are attributable to the "age" of the teacher in his or her career, his or her particular entry cohort (are teachers hired in the 1980s less committed, for example, than those hired in the 1960s), or the effects particular to that individual chronological year (1990).

Longitudinal data collection represents a first step towards unraveling these effects. If NCES collects longitudinal data on teachers who have already amassed varying years of experience in the base year of data collection, researchers can attempt to separately identify some of the effects by, for example, comparing the responses of teachers in their 6<sup>th</sup> year of service who were surveyed in the base year of the longitudinal study to those who were in their 6<sup>th</sup> year of service in a later year of longitudinal data collection.

But if all longitudinal data collection begins in the same base year, there will still be inseparable confounds (Baltes, ...) Moreover, NCES would have data describing the crucial first few years on the job only for one entry cohort (as subsequent data collection would focus on later years of the career). We therefore recommend that NCES commit to beginning the longitudinal cohorts in at least two and preferably three base years. In a later section, we will detail several possible alternative data collection schedules resulting from this recommendation.

**Design principle 6: Collect data from at all relevant levels in the organizational hierarchy**

Recognizing that teachers exist within an organizational context, NCES embedded the teacher interview component of the SASS in a series of linked surveys across several levels in the organizational hierarchy. Related questionnaires were sent to the LEA, the school principal, and in some cases, another school administrator. To keep costs at a minimum, the TFS component of SASS only surveys teachers, thus providing no organizational linkage for this second data collection point.

We recommend that NCES reconsider this omission in the design of a truly longitudinal study. We believe that it is a serious omission to cut the teacher data collection effort off from its organizational structure. With the current TFS, for example, we have no way of knowing to what type of district "movers" moved. Movers are asked a handful of questions, but it is difficult to determine whether this new district (or school) is different in fundamental ways from the teacher's previous district. Without such information it is difficult to build statistical models that might explain why the teachers left. (To see how difficult it is to build such models on the basis of available data, consult the Choy, et al., 1992.) Moreover, if NCES would like to use other respondents as informants about the target teachers, such multi-level data collection is imperative. By sending questionnaires to the principal of each sampled teacher's school, NCES might be able to collect a wealth of contextual data that might explicate the teacher's position in the school. Without linked questionnaires, researchers will have no way of corroborating the answers of the sampled teachers.

Not only do we recommend that NCES collect data on the organizational context of teachers as they move from school to school and district to district, we recommend that NCES also consider adding another level of data to the longitudinal study: data describing the *students* the teachers serve. One way of collecting such data would be to include additional explicit items on the teacher questionnaire, items asking teachers to describe the students in their charge in greater detail than is presently done. Another method which we raise here and recommend that NCES consider is to build explicit linkages between the longitudinal SASS and its other on-going NAEP data collection program.

Realistically, such linkage could only occur for a subset of teachers. Teachers will move schools, school districts, and states immediately after the base year of data collection. For the truly longitudinal study to accurately describe teachers' careers, *all teachers will need to be followed as they move jurisdictions*. Indeed, teachers who leave teaching will need to be followed out of the profession so that NCES can collect data on them as they reenter the schools. But even amongst those who stay, we expect a reasonable amount of mobility. Tracking these movers to their new schools

will represent a data collection challenge. But if it is possible to link only some of these teachers (even as they move) to NAEP data collection, this linked study would undoubtedly prove a vital resource to researchers in the years ahead.

#### Where do we stand?

These six design principles lie at the heart of our recommendation. We recognize that NCES may not be able to implement all of them when reconciling cost constraints against knowledge benefits. Some of these principles are clearly essential. NCES cannot, for example, conduct a rigorous longitudinal study without some form of prospective data collection. Yet it would be possible to study only a single base year cohort, recognizing from the outset the limitations this decision would create. So with these six principles in hand, we now turn to describing in greater detail our recommendations concerning the essential aspects of research design.

... to be continued

#### Selecting a target population: Whom should we study?

#### Time dimensions of data collection: For how long and how often should we collect data?

#### Measurement issues: How should the data be collected?

#### Next steps: Pilot studies

#### Conclusion

We believe that if the proposed longitudinal study adheres to these principles, it will represent a substantial leap forward from our current state of knowledge. Unlike studies based solely on administrative records, for example, this study would collect data from the teachers themselves. Unlike the analyses of national probability subsamples of teachers included in other large national surveys (such as the NLS-72 and the NLSY conducted by the Center for Human Services Research at Ohio State), this longitudinal study would gather data on a range of informants so that we may better understand teachers' worklives in context. Unlike the current SASS and TFS, this study would be truly longitudinal, not relying solely on retrospective reports of previous career decisions and limited one-year prospective reports. And to insure that we can begin to separate out the effects of entry year and year of the career, the study would track multiple cohorts, longitudinally over time.

#### References

- Billingsley, B. S. (in press). Teacher retention and attrition in special and general education: A critical review of the literature. *The Journal of Special Education*.
- Bobbitt, S. A., Faupel, E., & Burns, S. (1991). Characteristics of stayers, movers, and leavers: Results from the teacher followup survey, 1988-89. Washington, DC: National Center for Education Statistics.
- Choy, S. P., Medrich, E. A., Henke, R. R., & Bobbitt, S. A. (1992). *Schools and Staffing in the United States: A Statistical Profile*. Washington, DC: National Center for Education Statistics.
- Darling-Hammond, L. (1984). *Beyond the commission reports: The coming crisis in teaching*. Santa Monica, CA: Rand Corporation.
- Grissmer, D. W., & Kirby, S. N. (1987). *Teacher Attrition: The Uphill Climb to Staff the Nation's Schools*, Santa Monica, CA: Rand Corporation.
- Grissmer, D. W., & Kirby, S. N. (1992). *Patterns of Attrition Among Indiana Teachers: 1965-1987*, Santa Monica, CA: Rand Corporation.
- Haffner, A., & Owings, J. (1991). *Careers in teaching: Following members of the high school class of 1972 in and out of teaching*. Washington, DC: National Center for Education Statistics.
- Haggstrom, G. W., Darling-Hammond, L., & Grissmer, D. W. (1988). *Assessing Teacher Supply and Demand*, Santa Monica, CA: Rand Corporation.
- Heyns, B. (1988). Educational defectors: A first look at teacher attrition in the NLS-72, *Educational Researcher*, 17, 24-32.
- Murnane, R. J., Singer, J. D., & Willett, J. B. (1988). The career paths of teachers: Implications for teacher supply and methodological lessons for research, *Educational Researcher*, 17, 22-30.
- Murnane, R. J., Singer, J. D., & Willett, J. B. (1989). The Influences of Salaries and Opportunity Costs on Teachers' Career Choices: Evidence from North Carolina, *Harvard Educational Review*, 59, 325-346.
- Murnane, R. J., Singer, J. D., Willett, J. B., Kemple, J. J., & Olsen, R. J. (1991). *Who Will Teach?: Policies that Matter*. Cambridge, MA: Harvard University Press.
- Theobald, N. (1990). An examination of the influence of personal, professional, and school district characteristics on public school teacher retention. *Economics of Education Review*, 9, 241-250.
- Willett, J. B., & Singer, J. D. (1989). Two types of question about time: Methodological issues in the analysis of teacher career path data, *International Journal of Educational Research*, 13, 421-437.
- Willett, J. B., & Singer, J. D. (1991). From whether to when: New methods for studying student dropout and teacher attrition, *Review of Educational Research*, 61, 407-450.