

NCES IN THE IES ERA

American Educational Research Association

2004 Annual Meeting, Presidential Invited Session

April 14, 2004

by

Robert Lerner, Ph.D.

Commissioner, National Center for Education Statistics

Institute of Education Sciences

U.S. Department of Education

Introduction.

Thank you for giving me the opportunity to address the AERA on the subject of the role of the NCES in the IES era.

The Tasks of NCES.

NCES has been called “the jewel of the Department of Education,” and I am pleased to endorse that description. The collection, analysis, reporting, and dissemination of data, statistics, and reports on the “condition of education” continues to be an important duty of the Department of Education, and we are proud to be responsible for it.

In carrying out its duties, NCES has two primary points of reference. On the one hand, we are a federal statistical agency. We are a component of the federal statistical system along with our sister agencies such as the Census Bureau, the Bureau of Labor Statistics, and the National Center for Health Statistics. We are proud of the statistical standards we developed and published, which seek to insure the wide distribution of accurate data, while protecting respondent confidentiality. In continuing our mission, we strive to live up to the principles of the National Research Council report on federal statistical agencies.¹

On the other hand, NCES is now a major component of the Institute of Education Sciences (IES).² We are proud to be a part of this major effort in improving the rigor of education research. We participate by upgrading the rigor of our own work and utilizing

¹ See *Principles and Practices for a Federal Statistical Agency*. 2nd Edition. Washington, DC: National Academy Press, 2001.

² Whitehurst, Grover. “The Institute of Education Sciences: New Wine, New Bottles,” AERA Invited Paper, April 22, 2003.

the new possibilities for interdisciplinary cooperation made available by the enhanced IES.³

In carrying out its statutory duties, NCES serves the Department of Education and reports to Congress and the American people; and I take our reporting and statutory responsibilities quite seriously. These are not all of the constituencies we serve. Our data products and research reports are and should continue to be useful to academic and think tank researchers, to policy shops of widely varying ideological views, to education practitioners, and to members of the general public who are interested in education or have specific information needs concerning education. However, the above two points of reference, namely the federal statistical community and the Institute of Education Sciences, serve as the primary intellectual beacons by which we navigate our course in the seas of education data and research.

My View of the Education Sciences.

While there are many good things being done in the world of education research, there is much room for improvement. In a recent volume, the National Research Council concluded: “In no other field are personal experience and ideology so frequently relied on to make policy choices, and in no other field is the research base so inadequate and little used” (National Research Council, 1999, p. 1). I am not sure whether the research base is so inadequate because people see no need for it, or whether in fact people see no need for a research base because it is so inadequate. For those who prefer schematics, does A (a perception of no need for research) cause B (having an inadequate research base), does B (having an inadequate research base) cause A (a perception of no need for research), or

³ See also Mosteller and Boruch, 2002; National Research Council, 2002.

are A and B part of a vicious cycle, each reinforcing the other and producing neither a strong research base nor interest in improving the existing research base?

While I do not propose here and now to do the research necessary to find the answer to this question, I will hazard the guess that A (perception of no need for research) is the more important of the two causes. However, the fact that this is changing, as is evidenced by the formation of IES, brings hope that we can help improve the quality of the research base. Having done research in a number of social science fields, I am also not entirely sure that education research is so singular in its problems. However, it is true that much improvement is needed and it is also true that NCES as part of IES can play an important part in helping the field to progress.

The view of the education sciences I advance here is constrained by the fact that it concerns primarily what is relevant for NCES as a federal statistical agency. The research carried out by a federal statistical agency is not necessarily a reflection of the entirety of education research. This may be easier to see if we examine in some detail the kinds of questions asked by the social and behavioral sciences (SBS)⁴ in education and focus on those of special relevance to NCES.

Research Questions and Answers.

SBS research activities can be placed in three categories: 1) description, 2) association, and 3) explanation. Let me note at the outset that these are not judgmental

⁴ Exactly what to call the various disciplines is a problem. The label “social sciences” appears to omit psychology, which would be a great mistake, while the label “behavioral sciences” appears to omit economics and the relevant parts of political science, which would also be a great mistake. Thus, the label “social and behavioral sciences” (SBS) appears best.

categories, but rather as useful ways of describing the work to be done in education (and in other SBS fields).

DESCRIPTION.

What is Description?

Descriptive studies pose and answer the question of “what is that?” and seek to describe what is out there. A slightly more formal definition is: Descriptions consist of systematically recorded observations of what is out there according to systematic rules of classification.⁵ If one takes the IES mission to be the study of education and schooling throughout the life-course, then NCES’s descriptive data and studies based on them seek to describe selected aspects of the universe of educationally relevant “phenomena.”

The importance of accurate descriptive data.

In discussions of scientific research and methods, description is sometimes downplayed in favor of explanation. The former is sometimes considered to be trivial, while the latter is of sole importance. This is wrong. Of course adequate explanation in the form of scientific theory, containing a body of verified knowledge, is the ultimate goal of any scientific discipline, applied or theoretical. Experience in other scientific fields has shown that such development is the result of a cumulative effort requiring “conjectures and refutations” in a cycle of cumulative research that builds a growing body of knowledge. Contrary to its above reputation, descriptive data serve as building blocks for scientific theories and explanations. Properly understood and executed, careful description is a conceptual, even theoretical activity that requires careful attention to be

⁵ Adapted from Loether and McTavish, 1974, p. 14.

paid to questions of definition as well as to the process of data collection. No science can do without them, including SBS in education.

Data Collection and Dissemination at NCES.

Statistical description at NCES takes the form of collecting and disseminating descriptive data. This is a core activity of the kind of population-based studies that any federal statistical agency does and thus is a primary task of NCES. This has traditionally taken two forms: the first is exemplified in our universe collections--the Common Core of Data (CCD) and the Integrated Postsecondary Education Data System (IPEDS) databases. The second is exemplified by the descriptive components of our many surveys.

Our descriptive databases serve a multitude of functions. First, they provide basic publicly available information on schools and schooling at every level of the educational system: CCD for public schools including grades K-12, the parallel data collection effort called the Private School Survey (PSS) for private schools including grades K-12, and IPEDS for postsecondary education. These serve the public's need to locate basic information about particular institutions and, as important, serve NCES as sampling frames for our numerous surveys.

Our surveys are national probability samples of important populations, including education organizations. Respondents are asked detailed questions about schooling and educational topics and related information. In addition to the National Assessment of Educational Progress (NAEP), the "Nation's Report Card," these include the Schools and Staffing Survey (SASS), the National Education Longitudinal Study of 1988 (NELS: 88), the Education Longitudinal Study of 2002 (ELS: 2002), the Early Childhood

Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K), the Early Childhood Longitudinal Study-Birth Cohort (ECLS-B), the Beginning Postsecondary Study (BPS), the Baccalaureate and Beyond Longitudinal Study (B&B), our adult education surveys, and our international surveys. Many of the surveys are longitudinal data collections, structured to permit studies of changes over extended periods of time. For example, NELS: 88 begins with an 8th grade sample and follows them through college. These require considerable substantive and methodological skills to design, collect, analyze, disseminate, and report on. Developing and sustaining these research databases is an especially important task for a federal statistical agency such as NCES.

In general, ethnographies⁶ and documentary historical investigations, while they have their place in education research, are not an appropriate focus for an agency such as ours. Quantitative historical or trend analyses of student achievement has been done using both main and trend NAEP and by making comparisons among the National Longitudinal Study of the High School Class of 1972 (NLS-72), High School and Beyond (HS&B), and ELS:2002.

Collecting these data gives NCES substantial influence on what is studied in education and how rigorously it can be studied. This is because such data collection efforts determine what administrative records are collected and tabulated and what survey questions are posed to which kinds of respondents, the manner in which, and how often they are asked and tabulated.⁷

⁶ An exception might be the 1999 Third International Mathematics and Science Study-Repeat (TIMSS-R), which combines an international assessment with a video-recorded study of mathematics instruction in seven countries including the United States. (Hiebert et al., 2003). However, this is a comparative study using probability samples of classrooms in seven nations; it is not an impressionistic study by a single researcher. Also, academic achievement cannot be assessed qualitatively.

⁷ For more information on NCES surveys and activities, please see *Programs and Plans of the National Center for Education Statistics, 2003 Edition* (U.S. Department of Education, 2003).

As a result, we have to pay special attention to the quality and scope of the information we collect, regardless of whether it is obtained from administrative agencies or by questioning respondents (or very rarely, from direct observation). The point can be seen in a slightly different manner. Any well-trained researcher can compute regression equations using our data, but no one but NCES can collect the kind of data we collect as well we can.

Conceptualization and Measurement.

Since all description is to some extent concept-laden, proper description is inherently tied up with problems of conceptualization and measurement. Following from this, a second important NCES scientific task, one the agency is well equipped to pursue, is assisting in the proper explication of educational concepts and constructs.

In some respects this may appear strange, but it is not. As adults living in the society in which we work, we encounter as “natural objects” such persons as teachers, students, parents, principals, and administrators, and such institutional facts as classes, schools, school districts, and state education agencies as part of “everyday life.”⁸ This natural attitude (sometimes called a “common sense” view) serves us well enough in the business of daily life. However, it is inadequate for both scientific and public policy purposes, because these natural “typifications” are poorly conceptualized, poorly defined, poorly counted, poorly measured, and poorly related to each other.

⁸ The approach I have used draws from sociologists Berger and Kellner (1981), but it does not require accepting any particular theoretical framework and my views depart from their views on a number of points, including the utility of quantitative empirical data. An important contrasting framework also compatible with the approach taken here is given in Coleman (1990).

While considerable progress has been made in these activities, not nearly enough has been done to provide standardized definitions, measurements, and counts, so as to insure that researchers, policymakers, and the general public are all talking about the same persons, states, organizations, and processes. While descriptive accuracy by itself does not guarantee scientific progress, it is a necessary component of such progress.

In psychometrics, a field that is very relevant to education and education research, the importance of careful attention to problems of conceptualization and measurement is well recognized (e.g., Nunnally and Bernstein, 1994). In education itself, the need for the careful conceptualization and measurement of educational achievement is widely understood. Here one considers an underlying construct, say mathematics achievement, and one proceeds to measure it, often precisely enough to be used in individual assessment.⁹ For NCES, the development of IRT methodology and its utilization in NAEP and in our other sample surveys allow us to measure educational achievement of national random samples of 4th, 8th and 12th graders with a high degree of reliability, validity, and comparability over time. This represents a technical triumph of considerable magnitude. Even more of a triumph in this regard is the development of the marginal maximum likelihood (MML) methodology used to construct achievement test scores in NAEP and other NCES surveys. It is a matter of agency pride that such methodologies are being used in other countries and in other social science fields as well (e.g., measuring hunger and health).

We need to make similar progress in defining and measuring other important constructs in the fields of education and schooling, especially, but not only, predictor or

⁹ As a federal statistical agency, NCES does not perform individual assessment of the kind used in applied psychological work.

independent variables (see below). Proper conceptualization and measurement can include something as simple as ensuring that all school districts count the number of classroom teachers in each school, district, and state across the United States in the same way and at the same time.

It can also include counting or measuring something more complex, such as state or national high school graduation and completion rates. IES/NCES recently sponsored a conference on counting and measuring dropouts. It is not quite as simple as it might first appear. Some possible measures of school completion are: a status rate, an event rate, and a cohort rate among others. Which are better measures of this concept and for what purposes? How should GEDs and special state certifications for some kinds of disabled students be included (if at all) in measuring high school completion? Adding to the complexity, the No Child Left Behind Legislation (NCLB) has its own definition built into its authorizing legislation. The NCLB definition is:

“(... the percentage of students who graduate from secondary school with a regular diploma in the standard number of years).”¹⁰

To mention only two complications introduced: the phrase “regular diploma” would seem to exclude the General Educational Development (GED) test and the special certifications, while the phrase “in the standard number of years” would seem to imply that those who were retained in their grade would not be counted as graduates or perhaps be excluded entirely from the calculations (i.e., from the denominator as well as the numerator of any completion or graduation rate). Yet we know that many eventually graduate or receive GEDs.

¹⁰ 20 U.S.C. 6311(b)(2)(C)(vi).

There are many more such concepts to be defined with constructs that require precise measurement. Consider the following examples: charter schools, tracking and grouping, teacher quality, student engagement, classroom order/disorder, retention, homework, and parental involvement. The list is not meant to be exhaustive, only illustrative. These topics provide fodder for frequent disputes about educational matters among parents, professionals, policymakers, and researchers alike, but even a superficial scrutiny of these arguments reveals that there is little agreement on what is being talked about. This makes research difficult and coherent policy discussion and argument even more difficult.

Properly defined concepts can provide building blocks to construct a set of educational indicators both for the nation and for the states. Already, we have statistically reliable NAEP scores for every state in reading and mathematics. A more complete set of indicators would include measures of educational persistence and measures of supportive educational behaviors by parents and schools. It might also include measures of state behavior, such as the implementation of choice policies and of spending patterns. Again, this is not a complete list, but is meant to be suggestive only.

In sum, improving the conceptualization and measurement of other non-test score education variables is critical to the progress of education research and a key task that NCES can undertake because of its unique position in the world of educational research.

Statistical Association.¹¹

¹¹ The reader should be clear that by association, I mean any means whereby the relationship between two variables can be described. Associations can be expressed as mean differences, percentage differences, or correlations or as various kinds of ratios, including odds ratios.

As most descriptive data cannot be understood in isolation from other data or information, a second important NCES task is reporting on statistical associations among educational, schooling, and other variables. This takes two forms: first as agency reports and second as data that is made available for other researchers to use for their own purposes.

Many predictive factors are required or at least strongly recommended in our authorizing legislation as information we need to collect. We can and do routinely cross-classify educational phenomena by race/ethnicity, socio-economic status (SES), family structure, region, gender, disability, limited English proficiency, mobility, and urbanicity among others. These correlations are usually reported in the form of descriptive cross tabulations, applications of the general linear model, and other forms of multivariate analysis.

As with measures of individual indicators, these bivariate or multivariate descriptive analyses serve several purposes. First, they provide the findings that serve as the building blocks for theories about education. Well-known average educational and achievement differences among members of different demographic groups provide both a challenge to scientific understanding and a difficult policy challenge. The well-known Black-White test score gap is only one example of such a difference, as is the difference in achievement among individuals with different socio-economic backgrounds.

Second, they provide a useful set of factors that need to be taken into account routinely in order to construct more accurate scientific explanations of student achievement and other educational phenomena, even when the analyst is interested primarily in other issues because these differences are persistent, stable generalizations.

For example, comparisons of achievement levels among states are misleading unless these differences are taken into account.

Presenting data in the form of statistical associations serves the additional useful purpose of assisting in the development of a set of interrelated social indicators among ostensibly different phenomena, which may fall under different administrative purviews by an accident of administration. For example, it is likely to be the case that those students who do poorly in school and who behave badly there (e.g., cut classes, act out in class, fail to do homework) are also those who are likely to try drugs, engage in various kinds of delinquent behavior and the like, both in and outside of the immediate school environment.¹² Without presuming any particular view as to whether poor performance in school causes involvement in delinquent behavior, whether the reverse is true, or whether the statistical relationship is due to one or more common factors, it would be worth showing to what extent these behaviors tend to occur together as symptoms of underlying problems. This will be useful not only for researchers but for policymakers and even for practitioners and for parents, who can be alerted to the fact that one set of problems is likely to be symptomatic for all the rest and to act accordingly in their individual circumstances.

Explanation and Causation.

The payoff from all these activities is scientific explanation and its application to public policy issues. This requires the development of scientific theory based not only on descriptive generalization and laws but causal ones also. NCES does not and should not

¹² Hirschi and Gottfredson (1990) provide a detailed theoretical and empirical argument as to why this might be so.

have an “official” theory¹³ of educational achievement, but as noted above, the data that we collect and disseminate are gathered with certain assumptions in mind; we ask certain questions and not other questions because at some level these are concept-impregnated activities. For example, if we believe that parental attributes and behavior are important predictors of student achievement, we ask questions about parents or include parent questionnaires as components of our surveys. Hopefully, in doing so, we draw on the widest possible assortment of sources and ideas, so that we are not unduly limiting the scope of our data collection efforts.

Some Features of Research Design.

It would be nice to have highly developed theories of educational process, but these exist mostly in fragments (but see the interesting paper written by Walberg, 2003). Similarly, it would be nice to be able to tell policymakers and practitioners of exactly what works, but we often can't.

An especially important aspiration is to be able to make causal inferences. If we understand how educational phenomena behave, then the field would be able to use our findings to improve the outcomes of schooling and understand the operation of other educational processes (e.g., class size, school size, tracking, homework, private vs. public vouchers and the like).

In many instances, randomized experimentation is the most reliable way to make causal inferences. This is because experiments meet the classical criteria of ascertaining

¹³ Theories explain established empirical generalizations and predict as yet unknown empirical generalizations (Wallace, 1969, p. 90). A theory can be either deductive in that empirical generalizations are deduced from a small set of basic principles or they can be part of a pattern (e.g., Kaplan, 1964). There is no need here to discuss the enormous literature on the philosophy of science, but it does serve as background for what follows.

causality: correlation, time order, and lack of spuriousness, with the fewest assumptions. Random assignment by definition eliminates respondent selection. Controlling the administering of treatments to subjects automatically imposes a causal ordering on any results. Obtaining a statistically significant difference, then, provides strong evidence of causation. To put it in the language of Campbell and Stanley, such an inference has a high degree of internal validity.¹⁴

Non-experimental studies can also be used for this purpose, but it is much more difficult, especially to ascertain whether subtle differences among school factors act as potential causes for student achievement (or non-achievement). When causal order among the variables cannot be convincingly established, then the task is virtually impossible. When such order can be established, it requires detailed theoretical argument supported by extensive literature review, and above all, careful explication and testing of potential alternative explanations to eliminate possible sources of spuriousity. As might be imagined, this can be quite difficult to do.¹⁵

Here we reach a restriction on NCES's activities. Carrying out NCES's functions as a statistical agency well places limits on the kinds of studies NCES undertake. NCES's primary role as an agency is to provide and report on data that can be used to generate well-established empirical facts and generalizations that are critical to understanding the causes and dynamics of educational achievement and must be taken into account for any theory or explanation of educational phenomenon to work at all. With the exception of

¹⁴ See Campbell and Stanley (1966), Cook and Campbell (1979), and Shadish et al. (2002). Rossi et al. (2002) make a similar argument for evaluation research.

¹⁵ The case of smoking and lung cancer indicates that it is not impossible. Of course, this is predicated upon a number of unique conditions. First, the time ordering is easily established. Second, the statistical association between prolonged heavy smoking and the onset of lung cancer is very strong in conventional terms (odds ratios in the 20s). Third, many different kinds of research designs, each with different flaws, produced similar results. Many of these conditions do not apply in various parts of education research.

methodological experiments, such as those concerning response rates and different methods of survey administration, NCES does not perform or fund experiments; this is the role of other components of IES.

NCES does play an extremely important role as a critic of the overly simplistic explanations resorted to by others, because while it is hard to demonstrate causality by means of surveys, such data and evidence can often refute such claims.¹⁶ For example, individuals routinely attempt to infer causality from bivariate differences in test scores among states, without even taking well-established differences in population composition with different levels of achievement into account. This is an invalid inference.

I am willing to go even further than this by stating that without prior measures of student ability/achievement, no assertion of the putative educational impact of particular policies or programs that makes causal claims is likely to be worth very much. One implication of this is that reports based on cross-sectional surveys such as NAEP are unlikely to yield strong causal inferences as to the influence of many educational elements on academic achievement. A second implication is that our many longitudinal surveys, discussed above, are better avenues for exploring these issues. Such surveys provide one method for dealing with such problems in a survey research context.

The above discussion doesn't imply that randomized experiments are the perfect research method; no method is perfect for all purposes. In particular, it can be difficult to generalize from experimental results and experiments are often very expensive and slow to carry out. Also, there are some questions concerning education and schooling that cannot be answered by experiments for either practical or ethical reasons. For example,

¹⁶ For the sake of completeness, I note the possible existence of spurious non-correlation also exists, but this is less of a problem than spurious correlation.

research attempting to explain family and demographic influences on educational outcomes is unsuitable for experimentation. Neither can the effects of state policies such as high-stakes testing on achievement be studied in this way. Yet these are important subjects that can be studied in part by using NCES data.

We can highlight the NCES role in another way by referring to Shadish and Cook, and Campbell's (2002, pp. 37-102) classification of types of inferential validity of inferences in the context of various research designs: 1) internal validity, 2) external validity, 3) statistical conclusion validity, and 4) construct validity.¹⁷ While inferences using NCES data are normally weaker than experiments or quasi-experiments on internal validity, they are strong on external validity and representativeness, largely because of high response rates attained from probability samples of respondents. NCES research reports are also strong on statistical conclusion validity, because they use proper statistical procedures that produce correct standard errors and confidence intervals and because they use advanced statistical methods correctly. I have previously discussed the substantial construct validity of NCES measures of educational achievement and the importance of increasing the construct validity of other NCES educational indicators and predictor variables where they need improvement. Improving educational conceptualization and measurement of educational phenomena is the major focus of NCES as part of a scientific research agency.

EDUCATIONAL KNOWLEDGE AND CONTENT AS THE SUBSTANTIVE NCES FOCUS.

¹⁷ This phrase refers to whether an item measures what it is supposed to measure.

It is convenient to sketch out briefly some key foci for NCES data collection efforts by adapting a framework developed originally by a former professor of mine, James Coleman (1966 et al.; 1990; 1997 et al.). His original 1966 framework focused on academic achievement as the main outcome/dependent variable of schooling and he used both out-of-school and in-school factors as his main predictor/independent variables. I will use this framework to describe some problems and prospects in education research relevant to NCES.

Outcome/ Dependent Variables.

The primary outcome variables for NCES data collection activities should be the amount and growth of knowledge at all levels of education from pre-kindergarten through the Ph.D., and in the adult population. These include achievement test scores, grades (a much neglected set of indicators), completion/dropping out, retention, receiving educational honors, and the like. While psychometricians tend to prefer test scores to all other measures of academic achievement, a preference not without reason, grades continue to be given and are thus important outcomes to study. As far as I am aware, virtually every middle or high school uses grades of the familiar sort. Similarly, every college and university of which I am aware does the same. (Nota bene: This is not an argument for or against such practices. I merely state that since grades are widely used, they should be studied.)

Another kind of dependent variable specifically mentioned in our authorizing legislation is school crime and safety. NCES carries out two data collections in this area: 1) the School Crime Supplement to the National Crime Victimization Survey (a Bureau

of Justice Statistics household survey) and 2) the School Survey on Crime and Safety (SSOCS), a survey of principals. I hope to continue this kind of investigation but to expand the focus of such efforts to include school disciplinary issues and classroom disorder issues (e.g., Achilles, 2002, pp. 235-266). Of course, in the context of educational achievement, classroom order/disorder can be a predictor variable too.

While our data collections can and should be used by others for their own purposes, these are the key foci for NCES data collection efforts and reports. We are mandated to collect certain kinds of data (including some not listed here) and will continue to do so.

Predictor or Independent Variables.

There are two categories of variables that are of interest: school factor variables and non-school factor variables.¹⁸

Non-School Factor Variables.

Our data collections and descriptive reports should at a minimum contain the NCLB independent variable/predictor categories: race/ethnicity, gender, socio-economic status (low-income), limited English proficiency (LEP) status, and student disability (SD) status. They should also include such variables as urbanicity, region, mobility, family structure, and not least, measures of the value families and therefore students place on education and learning.

¹⁸ We are mindful that there are certain restrictions placed on what we can study under the provisions of NCLB.

They should include questions of parental control of student behavior, which is not as well understood as it should be. For example, what kinds of educational performances do parents find to be acceptable and what do they do about substandard performances? One study by Lawrence Steinberg et al. (1996, p. 161) described the minimum acceptable grades students were allowed to bring home and described what parent(s)/guardians felt about poor performance.¹⁹ We need to be able to replicate these descriptions, with the hope of eventually helping advise parents on how to better manage their kids.

School Factor Variables.

There are many important such variables for NCES data collection and reports of which I can list only a few here. Some of the relevant variables are organizational, such as the type of schooling attended: public (including charters), private, and home-schooled status. Some refer to units at higher levels of aggregation, such as the kind of external standards and testing schools are required to be accountable for. Studies of alternative schools are important also, as is the increasing use of ancillary services, summer school, after school programs, and tutorial services by parents.

Teachers (and professors) are an important focus of study. Obviously teachers matter in student achievement, but it is less clear exactly what about them matters. In the environments typical of today's schools, the quality of teaching may well matter more than any other single school-level variable in affecting achievement outcomes. There is

¹⁹ They reported that the average student-reported "trouble threshold," the lowest grade that could be received without parents becoming angry, was C- for Black and Latino students, for White students the "trouble threshold" was between a B and a C, and among Asian students the "trouble threshold" was A- (p. 161).

some empirical literature to support this view. One of the most striking findings of the original Coleman report was that teacher verbal ability seemed to matter more than other teacher and school characteristics in predicting student achievement; teacher ability had particularly noteworthy effects for Black students in predicting student achievement (Coleman et al., 1966, pp. 316-318).²⁰ This view was recently reaffirmed in two recent reviews of the literature (Mayer et al. 2001; Wayne and Youngs, 2003).²¹

There are a number of qualifications to be made to that statement. First, I do not forget my warning above about correlation not being causation. Although many of the studies cited in the above literature reviews use numerous control variables in presenting interpretations of their results, and the interpretations are plausible, there remains some question about selection effects. In other words, do high-ability teachers somehow manage to select largely high-ability students to teach? This requires further investigation.

Second, also following upon my methodological strictures above, I am unaware of any studies using randomized trials that assign high-ability teachers to an experimental group, low-ability teachers to a control group of students, and investigates the relative achievement gains of the two groups of students after some period of time has elapsed. Neither review cited above points to any such studies.

²⁰ Coleman subsequently stated that this was one of the most neglected parts of his original report (Coleman, 1990-1, pp. 18-20).

²¹ It is noteworthy that an extensive meta-analysis of the job performance literature suggests that for a wide variety of occupations over a long period of time, mental ability is a major predictor of job performance and trainability (Hunter and Schmidt, 1998).

Finally, teacher verbal ability alone is not likely to be the only important predictor of the quality of teaching.²² Obviously, teacher content knowledge matters also, especially in science and math.²³ We at NCES have been working on developing a paper-and-pencil set of measures of teacher quality that will encompass teacher content knowledge, but also teacher knowledge of student characteristics and teacher knowledge of optimal strategies for enhancing student engagement. This is very preliminary work, so I don't want to say much more now. I expect that we will eventually develop some useful measures of all these aspects of teacher quality.

Let me just briefly mention some other areas. The study of the number of courses taken in various subjects and the kinds of courses taken in these subjects (e.g., gifted and talented (GT), honors, advanced placement (AP), international baccalaureate (IB) courses, as well as specific subject matter courses) is a type of intermediate in-school outcome variable. On the one hand, it can be argued that the greater the rigor of a student's course selection and completion, the higher his or her level of achievement will be, at least up to some limiting point of course difficulty. On the other hand, this kind of investigation is subject to the methodological limits described above. The process of selection into such courses as influenced by student ability, parental involvement, and school guidance is itself a worthy subject of study, which can be carried out by using NCES data.

²² Hunter and Schmidt (1998) examine other factors that predict job performance, such as tests of integrity, conscientiousness, highly structured interviews, and work samples. The literature cited by these authors appears useful for suggestions on how to measure teacher quality.

²³ Stressed by Darling-Hammond and Youngs (2002).

Other important school policies include those mentioned above: tracking/grouping, retention/promotion, including social promotion, student engagement, classroom order/disorder, and school disciplinary policies.

In Summary: The Tasks of NCES

- NCES will focus on improving the conceptualization and measurement of relevant schooling and educational (dependent/outcome and independent/predictor) variables, so we can develop a system of education indicators that are usable for researchers, policy shops, practitioners, and the general public. In some cases, this may involve improving existing data systems and surveys, while in other cases it may involve collecting new data, either administrative or from surveys.
- NCES will continue to upgrade our statistical capacities, the capacity of our analysts, and our training efforts. This means we should learn to use the most advanced statistical tools and the most sophisticated means of carrying out our surveys.
- Other agencies have extensive programs for improving their sampling designs, question wording and the like. NCES needs to develop a coherent program of methodological research in this area. Such a program will lead to improvements in our data products and research reports.
- While the new IES structure has caused some trepidation, in fact, it represents an opportunity for the kind of interdisciplinary cooperation that could benefit all parts of IES. Education is an applied area of study and is thus inherently

interdisciplinary. While many at NCES are trained as sociologists, including myself, others in IES are trained in psychology. The potential synergy of such interdisciplinary collaboration has much to contribute to understanding education. We expect to utilize this potential to telling effect in the coming months and years.

In short, there is much work to do at NCES and I'm looking forward to rolling up my sleeves to work on these topics. Thank you. I welcome any comments and questions. My e-mail is robert.lerner@ed.gov.

BIBLIOGRAPHY

- Achilles, Charles M. "Racial Disparities in School Discipline," pp. 235-266 in *School Desegregation in the 21st Century*, Christine H. Rossell, David J. Armor, and Herbert J. Walberg (Eds.). Westport, CT: Praeger Publishers, 2002.
- Berger, Peter L., and Kellner, Hansfried. *Sociology Reinterpreted: An Essay on Method and Vocation*. Garden City, NY: Doubleday, 1981.
- Campbell, Donald T., and Stanley, Julian. *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally and Company, 1966.
- Coleman, J.S. "On the Self-Suppression of Academic Freedom," pp. 17-22 in *Academic Questions*, Vol. 4, No.1, Winter, 1990-1.
- Coleman, J.S. et al. *Equality of Educational Opportunity*. Office of Education/National Center for Education Statistics, Department of Health, Education, and Welfare, Washington, DC: Government Printing Office, 1966.
- Coleman, J.S. *Foundations of Social Theory*. Cambridge, MA: Harvard University Press, 1990.
- Coleman, James S., and Schneider, Barbara, et al. *Redesigning American Education*. Boulder, CO: Westview Press, 1997.
- Cook, Thomas D., and Campbell, Donald T. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago, IL: Rand McNally, 1979.
- Darling-Hammond, Linda, and Youngs, Peter. "Defining 'Highly Qualified Teachers': What Does 'Scientifically-Based Research' Actually Tell Us," pp. 13-21 in *Educational Researcher*, Vol. 31, No. 9, December, 2002.
- Hiebert, James et al. *Teaching Mathematics in Seven Countries: Results From the TIMSS 1999 Video Study*. Washington, DC, NCES 2003-013, Institute of Education Sciences, U.S. Department of Education, 2003.
- Hirschi, Travis and Gottfredson, Michael. *A General Theory of Crime*. Stanford, CA: Stanford University Press, 1990.
- Hunter, Frank L., and Schmidt, Peter. "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings," pp. 262-274 in *Psychological Bulletin*, Vol. 124, No. 2, 1998.
- Kaplan, Abraham. *The Conduct of Inquiry*. San Francisco, CA: Chandler Publishers, 1964.

Loether, Herman J., and McTavish, Donald G. *Descriptive Statistics for Sociologists: An Introduction*. Boston, MA: Allyn and Bacon, Inc., 1974.

Mayer, Daniel P. et al. *Monitoring School Quality: An Indicators Report*, NCES 2001-030R. Washington, DC: National Center for Education Statistics, U.S. Department of Education, 2001.

Mosteller, Frederick, and Boruch, Robert (Eds.). *Evidence Matters: Randomized Trials in Education Research*. Washington, DC: Brookings Press, 2002.

National Research Council. *Improving Student Learning: A Strategic Plan for Education Research and Its Utilization*. Washington, DC: National Academy Press, 1999.

National Research Council. Martin, Margaret, Straf, Miron L., and Citro, Constance (Eds.). *Principles and Practices for a Federal Statistical Agency*, 2nd Edition. Washington, DC: National Academy Press, 2001.

National Research Council. Shavelson, Richard J., and Lisa Towne (Eds.). *Scientific Research in Education*. Washington, DC: National Academy Press, 2002.

Nunnally, Jum C., and Bernstein, Ira H. *Psychometric Theory*. 3rd Edition. NY: McGraw-Hill, Inc., 1994.

Rossi, P. H., Freedman, H. E., and Lipsey, M.W. *Evaluation: A Systematic Approach*. 5th Edition, Thousand Oaks, CA: Sage Publications, 2002.

Shadish, William R., Cook, Thomas, D., and Campbell, Donald T. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin Company, 2002.

Steinberg, Lawrence, Brown, B. Bradford, and Dornbusch, Sanford M. *Beyond the Classroom: Why School Reform Has Failed and What Parents Need to Do*. NY: Simon and Schuster, 1996.

U.S. Department of Education, National Center for Education Statistics, *Programs and Plans of the National Center for Education Statistics, 2003 Edition*, (NCES 2004-027), Editor, William C. Sonnenberg. Washington, DC: 2003.

Walberg, Herbert J. "The National Assessment's Most Useful Background Items," unpublished paper, www.nagb.org, August 2003.

Wallace, Walter L. *The Logic of Science in Sociology*. NY: Aldine Publishers, 1969.

Wayne, Andrew J., and Youngs, Peter. "Teacher Characteristics and Student Achievement Gains: A Review," pp. 89-122 in *Review of Educational Research*, Vol. 73, No. 1 Spring, 2003.