DOCUMENT RESUME

ED 346 571 EA 023 997

AUTHOR Cooley, William W.
TITLE Confidentiality of Education Data and Data Access. Pennsylvania Educational Policy Studies Number 7.
INSTITUTION Pittsburgh Univ., Pa. Learning Research and Development Center.; Pittsburgh Univ., Pa. School of Education.
PUB DATE 30 Apr 90
NOTE 26p.
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Compliance (Legal); *Confidentiality; Confidential Records; *Data Collection; *Disclosure; Elementary Secondary Education; Ethics; *Privacy

ABSTRACT

Issues of confidentiality in educational data collection and access are examined in this paper, with attention to the treatment of confidential data. Confidentiality problems faced by the National Center for Education Statistics (NCES) are discussed in relation to the 1988 Hawkins-Stafford amendments, which protect tne privacy rights of individuals. The use of computers as threats to privacy and the importance of sharing data among researchers are also discussed. Suggestions for allowing access to NCES data include inventing an array of solutions, developing data files that prevent disclosure, and collapsing categories in descriptive fields. Data collection strategies are to develop uniform reporting procedures and to model state replications of relationships. Respecting individual privacy rights should be the first priority of the educational researcher. (13 references) (LMI)

EA

Pennsylvania Educational Policy Studies

PEPS is a joint effort of the U. of Pittsburgh's School of Education and the Learning Research and Development Center
This is policy paper number 7 in this series

Confidentiality of Education Data and Data Access

by

William W. Cooley

University of Pittsburgh

April 30, 1990

The purpose of this series of papers is to contribute to a more informed debate about critical policy issues facing Pennsylvania's public schools. This PEPS series draws upon a data base that has been established here at the University of Pittsburgh under the direction of William Cooley in cooperation with the Pennsylvania Department of Education.

Reactions can be shared:

by mail: LRDC, Pgh.,PA 15260
by PittVAX: COOLEY
by FAX: 412-624-7088
by PENN*LINK: PEPS
by phone: 412-624-7085
by INTERNET:COOLEY@VMS.CIS.PITT.EDU
by chat: room 743, LRDC

Nature of the Problem

The big challenge in a consideration of confidentiality of data and access to data by researchers is to reconcile the obvious need to understand what is happening in education today and the critical need to protect individual privacy and the confidentiality of individually identified data. The headlines today are about the need for better information about the condition of education and the need to better understand how education might be improved. Data about individuals--students, teachers, principals, parents--is a necessary ingredient if we are to improve that information and those understandings.

At the same time, it is imperative that procedures are in place which do not compromise individual rights, particularly the right to privacy. Researchers clearly must acquire a greater appreciation of the importance of individual privacy rights, and to establish procedures in their own work which reduce the dangers inherent in research that depends upon the collection and use of confidential data about people. Social scientists often try to argue that the problem is one of balancing the individual right to privacy and the social importance of their research. But the right to privacy is a fundamental right, and it is our responsibility as researchers to figure out ways of doing our research that do not violate that personal right.

There have been many examples of how data collections have become invasions of privacy, particularly when data were collected for one purpose and then used in ways which violate the conditions under which those data were collected. The report by the Secretary's Advisory Committee on

Automated Personal Data Systems compiled a good summary of such abuses (HEW, 1973), and the Privacy Act of 1974 was the culmination of their findings.

People have been worrying about the data collectors and record keepers for a long time. The Old Testament (II Samuel 24 and I Chronicles 21, 23, and 27) even provided an injunction against the census takers. Solzhenitsyn (1969) gave this graphic description:

> "As every man goes through life he fills in a number of forms for the record, each containing a number of questions...
>
> There are thus hundreds of little threads radiating from every man, millions of threads in all. If these threads were suddenly to become visible, the whole sky would look like a spider's web, and if they materialized as rubber bands, buses, trams and even people would all lose the ability to move...They are not visible, they are not material, but every man is constantly aware of their existence...Each man, permanently aware of his own invisible threads, naturally develops a respect for the people who manipulate the threads."

The law distinguishes four forms of invasion of privacy: 1) intrusion; 2) disclosure of confidential information; 3) publicly characterizing someone in a false or misleading manner; or 4) appropriating someone else's name or likeness for one's own benefit. This paper is mostly about the second form of invasion of privacy, the treatment of confidential data. It was written at the request of the National Center for Education Statistics (NCES), and so has as its focus the concerns about human privacy and confidential data that are

relevant to the sharing of the data that NCES collects.

National Center for Education Statistics

The purpose of NCES is "to collect, and analyze, and disseminate statistics and other data related to education in the United States and in other nations." (GEPA Section 406b) Toward that end, NCES collects vast amounts of data from individuals, usually with the provision that the data collected will be treated as confidential and will be reported only in statistical summaries that preclude the identification of anyone participating in the surveys. Some of the studies are cross-sectional, the most famous of which is the National Assessment of Educational Progress (NAEP). In these large, cross-sectional studies, only a minimum of identifying information is necessary in each individual record, the level of detail depending upon the sampling frame. For example, if sampling is done in a way the precludes generalizations below the state level, but estimates of state parameters are desired, then it is only necessary to know what state the individual resides in. Thus district, school, or individual identifiers need not be part of the student record.

But many of the NCES studies are longitudinal, requiring more detailed identifying information associated with each record so that individuals can be followed over time. In the most recent National Longitudinal Study (NLS-88), for example, the plan is to follow about 25,000 eight graders through high school and into their post high school careers, at two year intervals. Another major study, the Schools and Staffing Survey (SASS), seems to have begun as a cross-sectional study but has become a longitudinal study, with teachers,

for example, being asked to provide their home addresses for subsequent follow-up activity.

In the fall of 1989 a small controversy began when NCES felt obligated to delay the release of several reports pending confidentiality reviews. Report on Education Research (1989), for example, headlined "NCES Keeping Research Under Wraps Pending Confidentiality Reviews". The controversy confused the release of reports and the release of data because NCES has as its policy the release of data concurrent with the release of reports. The intent of that policy is to allow other researchers the opportunity to examine the data and determine if the conclusions made by NCES seem valid, and do so in a timely fashion. Thus the release of reports was held up while the associated data files were examined to see if confidentiality commitments would be violated by the release of the data.

These problems arose primarily because of the passage of the Hawkins-Stafford amendments of the GEPA in April of 1988, which greatly strengthened the nature and scope of NCES. Those new amendments also included a section (m) entitled "Confidential Treatment of Data", which stipulated the following:

"(4)(A) Except as provided in this section, no person may--

1. use any individually identifiable information furnished under the provisions of this section for any purpose other than the statistical purposes for which it is supplied;

2. make any publication whereby the data furnished by any particular person under this section can be identified; or

3. permit anyone other that the individuals authorized by

the Commissioner to examine the individual reports."

Another subsection makes it clear that the term 'report' means a response provided by or about an individual to an inquiry from the Center, but that this prohibition does not apply if the individuals identity cannot be revealed.

The bill also authorizes NCES to release tables and other statistical records to State and local officials, public and private organizations, and individuals, so long as confidentiality of persons is protected. Another critically important section stipulates that individually identifiable information is "immune from legal process, and shall not, without the consent of the individual concerned, be admitted as evidence or be used for any purpose in any action, suit or other judicial or administrative proceeding." [section (m)(4)(C)] Thus, if respondents are aware of this provision, they should be more willing to provide confidential information knowing that NCES could not be forced to reveal their identified responses to another government agency, for example.

It is very important to emphasize that many of the confidentiality difficulties which NCES currently faces are due to the fact that the 1988 amendments where passed while NCES was in the middle of several major data collecting operations. These experiences with trying to conform to that new law are suggesting many things to do differently in future data collections, both in terms of instrument design and in terms of data collection procedures.

In several conversations I had with NCES personnel and other researchers as I prepared this paper, the point was made that the concern about confidentiality derived from the 1988 amendments and <u>not</u> because

people had complained that their privacy had been violated. Caplan (1982) has an excellent chapter that relates to this point of "no complaints." If you assume, as Caplan does, that "without privacy it is not possible to develop or maintain a sense of self or personhood," then privacy is a basic human need. As such, a lack of concern about privacy or disclosure on the part of confidential respondents may be politically relevant, but it is not ethically relevant. Even if it were established that most people would not care if data about them were shared with researchers, this "would not constitute proof that we ought to loosen regulatory policies...Protecting the rights of the uninformed, the uninterested, or the incompetent may be paternalistic, but it is still morally important." (Caplan, 1982)

Definition of Key Terms

It is important that the reader know how I use various terms in this paper. I do not even pretend to be a lawyer, so I am not offering legal versions of such terms as "privacy". (In preparing this paper I have discovered how much lawyers disagree too!) These definitions are what I had in mind as I wrote this paper, however, so it is useful to know how I am using these words if you want to follow the various arguments and suggestions.

Privacy is the claim of individuals to determine for themselves when, who and to what extent individually identified data about them is communicated to or used by others. This includes the protection of an individual against harm or damage as a result of some record keeping operation, and against unwelcome, unfair, improper, or excessive collection

or dissemination of information, the intrusive nature of data collection, or of unwarranted data collection.

Confidential is a status accorded to data, and refers to how data, once collected, will be treated. Confidential status is usually determined by the conditions under which those data are collected. Confidential treatment means that anyone who has access to individually identified data is prevented from revealing that information to anyone outside of the immediate data collecting organization, or even anyone inside the organization who is not authorized to view confidential data.

Individual refers to any person, living or dead. A school, for example, is not an individual. A school's principal is.

A single record in a file of data about individuals contains information about a particular individual. Records consist of fields, e.g. a field for sex would indicate the individuals gender. Field values can be direct responses, "male", or can be coded using code keys (0=female, 1=male). Encrypted field values means that the data have been modified so that only a certain computer program could decipher the field values in a record. A secure code key means that the documentation for interpreting the values in a field are available only to authorized individuals.

An authorized individual is someone who has signed a nondisclosure affidavit.

Individually identified data are data which contain identifying information in one for more fields of an individual's record, such as name, social security number, phone number, or home address, which are more uniquely associated with that individual than, say, marital status would be.

Deductive disclosure means deducing who a record refers to even though those records have been stripped of identifying fields. A disclosure analysis involves establishing whether it would be possible for anyone to deduce the owner of a record by the unique characteristics of someone in that data file. "This record must be Mary Smith's since she was the only white, female, third grade teacher from the Lincoln School in the study."

It is also important to distinguish among three primary purposes for maintaining a system of records: administrative, security, and research (or statistical). The latter differs from the other two because in research, the individual's identity is not important, since the researcher is seeking generalizations across individuals, whereas for administrative and security purposes, decisions are being made about particular individuals, whose identity must be known. This paper is primarily about research-statistical files, although some consideration is given to the importance of being able to derive research files from administrative files.

Computers and Privacy

Although people have worried for centuries about the ways in which records being kept about them could become an invasion of their privacy, people really became concerned with the advent of modern data processing. As a result of these concerns, much was written in the late 60's and early 70's on how to deal with the new computer revolution. This activity culminated in the federal Privacy Act of 1974, the first serious effort to come to grips with the threats to privacy created by the new electronic age.

An excellent summary of what was known and thought about that

problem just prior to the passage of that 1974 act is in the report of the HEW Secretary's Advisory Committee on Automated Personal Data Systems entitled Records, Computers and the Rights of Citizens (HEW, 1973). Their introduction to the history of this problem showed how the notion of a research database or "statistical file" evolved:

> "The problem of gathering information from an antagonistic public led to the creation of yet another class of official records, the so-called statistical file. The essence of such a file is that the data it contains are not used to affect specific individuals. In creating such a file, the government, in order to gain information the public might otherwise be reluctant to give, forgoes some of the power over individuals that administrative records containing the same data would afford. The essential condition is that citizens believe that their individual contributions to a statistical file will not be made public and will not be used to punish or embarrass them."

An important aspect of computer based data files that is often overlooked, given all the concern about the ease with which computer files can be searched for particular records, is the ease with which computer based data files can be stripped of identifying information. In contrast, the traditional document ("paper") file of individual records that contain identifying information on each record cannot easily be made anonymous. It is probable that such physical files can also be penetrated by unauthorized individuals more easily than can computer based data files. This needs to be recognized and taken advantage of as we work toward policies and procedures for protecting people from invasion of their privacy.

National polls have shown that a majority of people believe that computers are a threat to privacy. It seems imperative that those of us who have come to appreciate the social benefits of computer based data files work hard to develop the procedures and technical solutions necessary to reassure the public that this is not necessarily so.

Some Biases I Bring to this Paper

It is important to reveal the perspective from which I approached the task of preparing this paper. I am an educational researcher who has worked with confidential data from students, teachers and others since 1958, when I launched the Scientific Careers Study, an overlapping longitudinal study of 800 young men who appeared to be heading toward careers in science. This was my first lesson on how much one can learn when people are willing to share confidential data with researchers. (Cooley, 1963)

Next, I directed Project TALENT, which was the first national longitudinal study of American youth, a 5% sample of 400,000 students in grades 9 to 12, tested in high school in 1960, with follow-up surveys conducted at particular points in their career development. As is usually the case in federally funded studies in education, the resources available to analyze this vast and expensive data collection was nowhere near the potential, so we established the Project TALENT data bank (Cooley, 1965), and defined procedures so that others could gain access to these data. Because the data files were vast and complex, and because of confidentiality commitments we had made to students, teachers and administrators, we did not share raw data, but rather conducted analyses for data bank "customers" using the following

procedures:

"1. The researcher sends a request to Project TALENT Data Bank Coordinator at the University of Pittsburgh.

2. The request is given a preliminary screening by the Data Bank Coordinator at Project TALENT.

3. The Project TALENT staff will meet to make recommendations on the action to be taken on each research proposal.

4. Time and cost estimates are sent to those researchers whose requests are approved.

5. Analyses are performed upon receipt of "OK" from initiating researcher.

6. Results are sent to the researcher for interpretation."

(Cooley, 1965)

Although this became an active data bank at the time, there was one major disadvantage--external researchers could not really "play" with the data. As data analysts know, much can be learned about the data from exploratory manipulations which are very difficult, if not impossible, to lay out in detail in advance. We did not share data because we did not have the resources (or, quite frankly, the expertise) needed to prepare data files that would not be a violation of our confidentiality commitments. So the above procedures were the best we could do, but were not completely satisfactory.

My next experience with confidential data was between 1978 and 1984 when I was doing evaluation and policy studies for the Pittsburgh Public Schools. At that time the district had no research office, so as we were

asked to do more and more studies, we built a rather impressive database, which included detailed data on each student in the district. Agreements were made and procedures were instituted which protected the privacy of individuals in the database. We were essentially the official research office for the District, and we returned all data and tapes to the District when it established its own in-house research office.

In my current work I have established a database which includes data on the students, teachers, schools and school districts in Pennsylvania. The Pennsylvania Department of Education shares these computer based data files with me because I have offered to conduct studies for the educational policy makers in Pennsylvania. I have signed a letter of agreement indicating that I would never reveal data or results that could be linked to specific individuals. All of the graduate students and faculty colleagues who access this database also agree to abide by that condition. But rather than rely on everyone's good will, we simply do not have any individually identified data in the research database.

Because the research requires our ability to link data across schools and school districts, a numerical code for school and district is in each individual record. In the research data base, we will be using a secure code key for school, since it would be possible to find a particular teacher's record, for example, if someone came to the database with lots of information to make such deduction, such as knowing the age, sex, race and birthday of a given teacher in a given school. It is, of course, unlikely that anyone would go to all that trouble since they would not learn anything from that teacher's record that was not already publicly available.

However, many schools ha only one teacher at, say, third grade. So, in our total database it would be possible to identify that school in the state with the lowest performing third graders on a state math test. Making that generally known could be harmful to that teacher because there are some people out there who might clamor for that teacher's resignation, even though there may be no justification for the causal inference thus implied. That would be an invasion of privacy in the sense that it would be publicly characterizing someone in a false or misleading manner.

Data consisting of individual records are critical to a researcher who wants to examine relationships and aggregations across individuals, but individually identified data are not required for those research purposes. But simply removing the individually identified data from each record is insufficient, as long as deductive disclosure is possible. The trick is to eliminate the possibility of deductive disclosure.

When longitudinal data are required (i.e. individual data linked over time), those linkages are established within PDE's computer, and the linked records stripped of individually identified data are shared with us. So as far as individually identified data is concerned, we have none, we want none, we need none, but as a precaution, everyone accessing the data agrees not to report any results that could be linked to a particular individual or school.

Importance of Sharing Data Among Researchers

For centuries researchers have benefitted from data that were collected by other researchers. This is especially true where systematic observations over time are crucial. Kepler would not have been able to deduce his three

laws of planetary motion if it had not been for Tycho Brahe's database. It was the carefully recorded observations of hundreds of botanists, zoologists, and geologists which made it possible for Darwin and others to piece together a convincing picture of natural selection.

More recently, social scientists have come to appreciate the importance of having longitudinal databases available for their work, but unlike stars and plants, legitimate concerns about the invasion of privacy make the task more difficult. However, technical solutions are clearly possible, and more work needs to be done to improve these techniques. This includes developing procedures for creating research data files from administrative files. The latter is important to reduce the need for data to be collected solely for research purposes, which is both expensive, and which adds to the intrusions into peoples privacy.

My current state policy work has convinced me that we have greatly underestimated what can be learned about education from the data which states are mandated to collect as part of their normal operations. These administrative data, being part of the public record of governmental organizations, can be shared with researchers because they are not confidential data. However, as is pointed out elsewhere in this paper, it is possible to organize and publicize such data in ways that become an invasion of privacy, but not if the data files are constructed in a way that eliminates individually identified data and guards against deductive disclosure.

Data collection as an invasion of privacy

It is also important to protect individuals against unwelcome, unfair, improper, or excessive collection of data, or data collection that is intrusive in nature, or of unwarranted data collection. This concern, of course, is one reason why OMB has been charged with the control of data collection through their clearance procedures. So the more we can develop procedures for sharing data in ways that protect privacy rights, the less need their is to collect the same data for different purposes.

Another important reason for further developing the capability of data sharing is the need for timeliness in policy studies. Making administrative data available for research purposes can greatly reduce the time and effort required to inform current policy issues. People on the receiving end of the data collector's queries know first hand how often it is that they provide the same information to different bureaus within the same government agency. By forcing the sharing of data which different bureaucrats collect we can greatly increase the potential for timeliness in educational policy studies.

Another reason why it is important to improve our capacity for sharing data among researchers is the advantage of multiple perspectives in the analysis and interpretation of research data. Policy studies deal with political issues. The unbiased researcher is a myth. If policy research is going to improve the quality of the debates as politicians resolve policy issues, it is critical to have researchers "keep each other honest" through multiple looks at the data.

Some people try to draw a distinction between data that are "sensitive" or not sensitive. That does not seem to me to be a useful distinction. There

was a time when my age and my weight were not sensitive data, but not today. What is important is the condition under which those data were collected. If I am asked to fill out a questionnaire and told that the answers I provide will be treated as confidential, then all the information that I provide, regardless of whether some subsequent researcher decides that some of that information is not sensitive, must be treated as confidential. This means that my name could never be publicly associated with my responses.

Similarly, if someone provides confidential information to their employer that is necessary to their employment, that information cannot be shared with other organizations in ways that makes it possible for that confidential information to be identified with the individual who provided that information. But such information could be shared if deductive disclosure were not possible.

But another consideration here is the notion that some data about individuals are publicly available because they work for public institutions, such as public schools. Releasing data that made it possible for someone to deduce that the salary of the principal of the Lincoln School was $40,000 last year is not an invasion of her privacy since that information is publicly available. If, however, in a confidential Schools and Staffing Survey, that principal reveals that a serious problem in her school is the physical abuse of teachers, and in the subsequent public release of a data file it would be possible to identify her record because someone knew she was a respondent in that study and she was the only one in that data file who had a masters degree in law, and someone also knew that, then that would be an invasion of that principal's privacy, whether or not that someone told anyone else their

deduction.

It is not an invasion of privacy if researchers gain access to data files in which the individual cannot be identified. An identifiable, confidential record cannot be shared with anyone who lies outside of the scope of the assurances given when those data were collected. But if a data file has been subjected to an adequate disclosure analysis, and the implications of that analysis have been implemented, then those records can be shared. In terms of invasion of privacy, there is no difference between publishing a statistical table that reveals that some of the respondents to this survey are members of the communist party, and releasing a data file that reveals that some of the individuals in the file belong to the communist party.

Allowing Access to NCES Data

To meet the legitimate needs of educational research, while protecting the individual's rights to privacy and the confidentiality of data, it is necessary for NCES to adhere to definite policies and procedures for the release of data files. Some researchers try to argue (e.g. Wallace, 1982) that the social importance of their research is sufficient justification for overriding the individual right to privacy, but as Pinkard (1982) argues, there is a fundamental human right to privacy. Social scientists do not have a right to override that more fundamental right. This right to privacy is violated when confidential information is released to the public in ways that make it possible for individual records to be identified.

It is quite insufficient to assume that researchers are "good guys." Of course, "I'm a good guy", but you can't trust everyone. It is very important

to rely more on technical solutions than nondisclosure affidavits, but the latter are clearly also important, for it encourages "goodness" where technical solutions are not possible. Boruch (1982), after reminding the reader that "Mark Twain defined an ethical man as a Christian holding four aces", dedicated his paper "to providing decent cards, if not aces, to the researcher who would be ethical." He then went on to provide the best summary I could find for developing procedural and technical solutions to the educational researchers privacy problems, both for data collecting and for data sharing. Boruch (1969, 1982) seems to be one of the few researchers who has emphasized the need for research on this problem of procedural and technical solutions to prevent deductive disclosure.

Since Boruch's papers are readily available to NCES, and since specific solutions tend to be context dependent, the details of his suggestions need not be repeated here, but the implications for NCES of his general approach is important to summarize. The first implication is that NCES must invent an array of solutions to its confidentiality problems. Such problems tend not to have a unitary character. Different kinds of surveys, or the building of NCES's common core of data, may require different approaches.

A second implication is that NCES should not rely on wide-spread oath taking as the preferred solution. Procedural and technical solutions may be more costly, but they should be developed as a high priority. Staff who deal with individually identified records, or conduct the disclosure analyses and prepare data files in which deductive disclosure is not possible, must be sworn to observe the confidentiality of the data. But beyond them, the emphasis should be on the development of data files and reports which

prevent such disclosure.

Finally, NCES needs to distinguish its real confidentiality problems from "mistakes and red herrings". The 1988 amendments forced a greater awareness of these problems, and it is important that this be viewed as an opportunity to work toward concrete procedural and technical solutions, as opposed to political or theoretical or oath taking solutions. There seems to be no big, immediate need to change the law that was just passed. What needs to be done is to examine what kinds of specific problems the procedures used to date have created and then work to invent solutions that would reduce the likelihood of their reoccurrence.

The dangers of deductive disclosure can be greatly reduced by having fewer categories in descriptive fields. Looking at the SASS surveys, for example, much could be done in collapsing categories without loss of policy relevant information. Such collapsing should be done from the beginning. For example, there seems to be no good reason to have 84 categories for major field code when teachers are asked to describe their academic background. That type of detail could easily cause uneasiness in the respondent, is more detail than any policy issue would require, and would difficult to interpret in any relational (cross-tabulation) explorations using all 84 categories.

NCES is in an excellent position to make some important contributions to the technology of disclosure-avoidance. It is necessary for them to do so. They are obligated by law to protect the confidentiality of subjects, and at the same time obligated to share data with the research community. If "necessity is the mother of invention," then NCES is very pregnant. They also have the

talent now to develop procedures and techniques for preventing deductive disclosure.

As Caplan (1982) pointed out, humanists sometimes rebel at the notion that there may be technical solutions for solving ethical problems. But surely it is important to establish how we can minimize the risks of invasion of privacy as we try to optimize our ability to improve education in the United States.

Of course the identifying information needed for longitudinal tracking of students need not be part of the data files used by people who are analyzing these data. NCES has already partitioned off such identifying files and the linking strategies necessary for longitudinal tracking. Only authorized individuals have access to such files. These are well established procedures and techniques.

In terms of data collecting strategies, NCES has already recognized the need to obtain disclosure affidavits from field coordinators in the future. The problem has been that the field coordinators know who the respondents were and as a result might be able to figure out the identity of one or more records in a subsequently released data tape. But it would be important to consider whether the types of procedures which Boruch suggest would be an even more effective way of dealing with this problem.

National Cooperative Education Statistics System

An important component of the Hawkins-Stafford 1988 amendments was the establishment within NCES of the National Cooperative Education Statistics System. As the bill points out in a new subsection (h), "The purpose of the System is to produce and maintain, with the cooperation of the

States, comparable and uniform educational information and data that are useful for policy making at the Federal, State, and local level." This amendment recognizes the potential for deriving much useful educational data from the current data collecting operations that take place within states and LEA's.

In my current work in Pennsylvania I have become quite impressed with what can be learned about education in Pennsylvania with data that are collected as part of state and local operations. This was in part possible because the state has developed uniform reporting procedures for the 501 school districts in the state. This makes it possible to look across districts for generalizations about relationships and across time for longitudinal trends. All of this is possible without collecting any new data and without endangering peoples privacy rights.

The bill quite properly emphasizes the need for further developing the common core of data that are available through NCES for all states, districts and schools. This requires agreement on definitions and procedures among the states so that such phenomena as dropping out of school can be studied across states. However, it should also be pointed out that much can be learned from within state replications of relationship seeking even if particular indicators are not on the same scale. Improving our ability to model educational phenomena within state could have important national implications.

For example, most states now have mandated, state-wide testing programs. However, different tests are given at different grades at different times for different reasons in 50 different states. Trying to get states to do

this testing in ways that would allow direct comparisons of student achievement among states is not feasible, even if it where considered desirable. Nor would it be feasible to expand NAEP to accomplish such a goal. But it would be possible for NCES, under the 1988 amendments, to provide technical assistance to the states so that much more can be learned about the dynamic relationships that exist within states among indicators of such domains as student performance, student demographics, teacher characteristics, and expenditures and revenues. Such an effort would greatly enhance our understanding of the current condition of education, much more than has the comparison of states on "off the wall" indicators such as ACT and SAT scores.

Conclusions

In the preparation of this paper I have become impressed with what NCES has become. Under Emerson Elliott's leadership the Center has attracted an impressive group of people and are doing much more data collecting and sharing than I was aware of. As I read about privacy problems and talked with colleagues in the field (the recent meeting of AERA provided that opportunity), it seemed to me that it was more important to educate researchers about the need to protect confidential data than it was to educate NCES staff on how to do it.

Many of our research colleagues became impatient with the delays in releasing data which occurred following the 1988 amendments. It now seems to me that those delays were necessary, and if NCES learns from those experiences, and I believe they are, we will have access to data that are useful in improving our understanding of how education works, and will not

violate individual privacy rights. The later must be the first priority.

REFERENCES

Beauchamp, T.L., Faden, R.R., Wallace, R.J. Walters, L. (editors) Ethical Issues in Social Science Research, John Hopkins University Press, 1982.

Boruch, R.F. Educational Research and Confidentiality of Data, American Council on Education Research Reports, 1969.

Boruch, R.F. "Methods of Resolving Privacy Problems in Social Research" in Beauchamp et.al., 1982.

Caplan, A.L. "On Privacy and Confidentiality in Social Science Research", in Beauchamp, et al, 1982.

Cooley, W.W. The Career Development of Scientists: An Overlapping Longitudinal Study of Scientific Careers, Harvard Graduate School of Education, 1963.

Cooley, W.W. "The Project TALENT Data Bank" Journal of Educational Measurement, Vol. 2, No. 2, December 1965.

Hawkins-Stafford Amendments to the General Education Provisions Act (GEPA), April 1988.

HEW Secretary's Advisory Committee on Automated Personal Data Systems Records, Computers and the Rights of Citizens, MIT, 1973.

Knerr, C. Confidentiality of Research and Statistical Data, Law Enforcement Assistance Administration, U. S. Department of Justice, 1978.

National Research Council Privacy and Confidentiality as Factors in Survey Response National Academy of Sciences, 1979.

Pinkard, T. "Invasions of Privacy in Social Science Research" in Beauchamp, et. al. 1982.

Report on Education Research, "NCES Keeping Research Under Wraps Pending Confidentiality Reviews", Capital Publications, Vol. 21, No. 18, September 6, 1989.

Solzhenitsyn, A. Cancer Ward, Rarrar. Straus and Giroux, Inc., 1969, as quoted in HEW (1973).