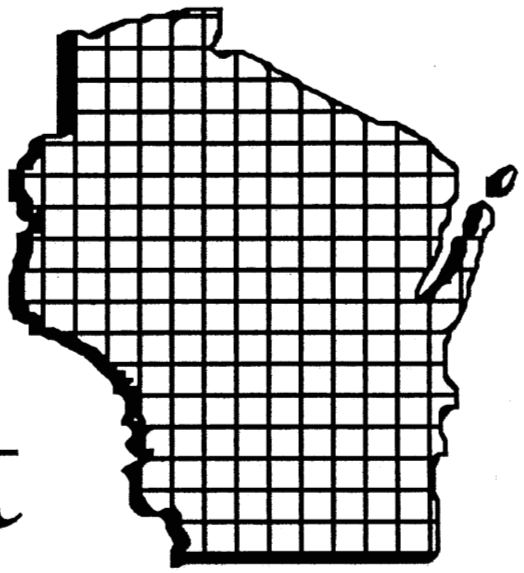
Wisconsin

Policy Research Institute Report



December 1989 Volume 2, No. 9

EVALUATING WISCONSIN'S TEACHERS: IT CAN BE DONE

Report from the Executive Director:

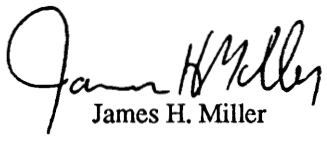
In recent public opinion polls by the Wisconsin Policy Research Institute and Wisconsin Education Association Council, the majority of Wisconsin residents believed that there should be merit pay for teachers. The problem has been how can you measure teacher performance in schools.

Whenever this issue has been raised, the criticism has been that teacher performance can only be judged by qualitative means. This new study develops a method that judges teachers performance quantitatively. The study was done by Professors Richard Bingham of Cleveland State University and John Heywood of the University of Wisconsin-Milwaukee. Bingham has a national reputation in Urban Affairs and Heywood has a strong background in statistics.

What is interesting about their method is that it can be used fairly to judge teachers working with disadvantaged students in the inner city or with teachers working with suburban or rural students. In their study they analyze the results from over 100 elementary schools using 5th grade students as their base. The idea was that each year as a student begins 5th grade there are academic expectations that can be predicted for the student. At the end of that grade one presumes that the student's performance reaching or not reaching their expectations is influenced by the role of the teacher.

Their data deals with a number of demographic factors that influence the student's performance. While this method is a new way of dealing with teaching performance, it must be pointed out that judgements can only be made over several years based on constant quality data.

The purpose of this study is to find a fair basis of rewarding competent teachers with additional pay. In addition, teachers who do not perform well can be helped to improve their performance by those teachers who do perform well. This idea appears to be one way of answering the age old question of whether or not teachers can be graded. The answer is yes.



James H. Miller

THE WISCONSIN POLICY RESEARCH INSTITUTE

3107 North Shepard Avenue
Milwaukee, WI 53211
(414) 963-0600

EVALUATING WISCONSIN'S TEACHERS: IT CAN BE DONE

by

Bingham, Heywood

TABLE OF CONTENTS

EVALUATING WISCONSIN'S TEACHERS:

IT CAN BE DONE

by

Richard D. Bingham, Ph.D.

John S. Heywood, Ph.D.

EXECUTIVE SUMMARY

The growing concern about the quality of education being provided by our nation's public schools is being shared by citizens and public officials alike. Increasing evidence shows not only that American students are learning much less than their counterparts in other industrialized nations, but also that today's American students, on average, are learning considerably less than their counterparts of twenty-five years ago. Hard questions are being asked about who is responsible for these sorry conditions and what can be done about them.

Unfortunately, we do not have irrefutable answers to either question. But prominent in proffered answers to each are teachers and improved teaching. It is teachers who spend large amounts of time with students and who many believe should have the greatest impact on student performance. But how can teachers be held accountable for student performance when to date we have no accurate way of measuring their contribution to student learning? And how can we improve teaching when we cannot prove what are the superior ways to teach?

This report describes a newly developed method which attempts to answer these questions. It outlines a systematic means of identifying teachers whose students annually have learned more than they were expected to learn, based on their previous performance. Superior teachers are identified through a procedure which involves statistically predicting how each student in a classroom should do in a given year based on previous performance and comparing the prediction with actual performance. The residuals (differences) between the two for each student are then aggregated across all students in the classroom to determine whether students as a whole did better, a positive sum, or worse, a negative sum, than predicted. Superior teachers are identified when students did much better than predicted and thus have a large positive residual. Inferior teachers are identified when students did much worse, as shown by a large negative residual for the classroom. A similar procedure is used to identify superior schools.

In the pilot project in which this study was undertaken, some extreme boundaries were set for identifying superior and inferior teachers and schools. The boundaries were arbitrarily set to prove that there is a continuum of results that would identify superior to inferior schools and teachers. Out of the almost 100 schools in the sample, seven schools had unusually large, positive results, and ten had unusually large, negative residuals. At the classroom level eight teachers were associated with unusually positive and twenty-four with unusually negative differences.

The most important conclusion is that one can begin to objectively evaluate teachers using this method. The method is flexible enough that it can be customized for any school district. It provides an identification of the superior and inferior teachers and holds teachers accountable for the performance of their students, taking into account other factors affecting performance.

Once the superior teachers have been identified, perhaps further answers as to how to improve teaching can be gleaned. The mere fact that teachers can be held accountable may help. But by identifying the superior teachers and then monitoring them, precisely what it is that makes them successful can be learned and passed on to others. Second, those superior teachers who desire it might be elevated to master teacher status so that they can directly convey their methods of teaching to others. Third, the successful teachers should be financially rewarded for achievement, creating incentives for them and others. This pilot method allows merit pay plans to be implemented because it identifies those teachers who are most deserving.

On the negative side, further testing and/or implementation of these procedures does impose certain costs. Complete and accurate records on each student must be maintained in a centralized, computerized file. Without such a record-keeping system the procedure should not be attempted.

MEASURING TEACHER PERFORMANCE

There is increasing concern with the quality of education being provided in our public schools. At the national level, numerous studies and reports have decried the apparent decline in the intellectual performance of American students over time. Among the most infamous is the 1983 report of the National Commission on Excellence in Education entitled, A Nation at Risk: The Imperative for Educational Reform, which severely criticized the quality of American education. Recently our current Secretary of Education, Lauro F. Cavazos, assessed the changes in education since the 1983 report and observed that only marginal improvements had been made.

At the state level there is parallel concern for the quality of education. In a state which has always prided itself on the quality of its educational institutions, recent surveys have revealed a deep-seated citizen discontent with the Wisconsin schools.[1]

If, as many do, we conclude that many young citizens are less well prepared than their elders to accept the responsibilities thrust upon them by life in a post-industrial democratic society, the problem assumes more gravity when comparisons of the achievement and training of American youth are made with comparable populations in other, similar societies. American educational achievement levels are below those in many industrialized, competitor nations.

Given wide consensus on the existence of a crisis in public education and the urgent need for reform, old questions have reappeared, questions that have been at the forefront of educational research for many decades. Perhaps the most salient are: "Who is responsible for the state of learning in American society?" and "How can the education of our youth be improved?" These questions raise the issue of accountability and effectiveness in the educational process.

Many persons and institutions are currently being blamed for the sorry state of learning in our primary and secondary schools. Teachers are an obvious target, since it is their job to educate. But others being fingered are the students themselves, student background, students' parents, school boards, school administrators, teachers' unions, and others who may play some role in education. It is difficult to precisely place the blame because strong cases can be and have been made for each. No one has resolved this, nor can we. Instead we will focus on a high profile group, teachers. It is they who spend the most time with the students and who many believe should have the greatest impact on student performance. Teachers may not be solely responsible for student achievement, but nearly all will agree that they do play an important role.

The issue is one of "accountability." This is a report on an empirical investigation designed to determine if it is possible to hold individual teachers accountable for the academic performance of their students.

WHY MAKE TEACHERS ACCOUNTABLE?

Why should anyone be concerned with making teachers accountable for the performance of their students? The most obvious reason is that if this link can be made, it would be possible to have a positive impact on the quality of teaching and on the subsequent performance of students. If we can determine what impact teachers have, we can identify the better and the worse teachers. By identifying the better teachers, three benefits could eventually be derived: 1) the teaching techniques of the better teachers could be studied and passed on to others so that more teachers could emulate their style; 2) better teachers could be offered legitimate positions as master teachers to help others; 3) better teachers could be financially rewarded on the basis of an objective evaluation system, thereby overcoming the major barriers which exist today for merit pay plans. By identifying the weaker teachers, it would create the opportunity to provide additional assistance and counseling, as they now do in Rochester, New York, to those teachers who are struggling the most. Just by creating and implementing a system that objectively evaluates teachers, teachers themselves would benefit because they would find out how well what they do actually works.

RECENT ATTEMPTS

Numerous efforts have been tried as means to increase student performance through changes in teachers. The changes have been based on various theories holding that teachers have not been effective. The theories range from the charge that the quality of teaching has deteriorated over the last twenty-five years, that what it is that teachers should be doing to be successful is not well known, that teachers know what to do but are just not doing it, to the assertion that teachers do not have sufficient power to be effective regardless of what they know. Each of these theories will be discussed, but the bottom line is that by many measures a number of students are still not learning at the rate we expect them to.

One theory as to why student performance has deteriorated over the last two decades is that there has been a significant decline in the quality of those persons attracted to the teaching profession. One source states that, based on an analysis of SAT results, those seeking a career in teaching were disproportionately drawn from those with the lowest test scores, that the average absolute score on the Verbal Ability portion of the test was substantially below the average of all taking it, and that scores of those interested in teaching averaged lower than all other fields save ethnic studies.[2] There is not a necessary relationship between high test scores and ability to teach, but many persons prefer that their children be taught by other than the least capable college graduates.

One of the reasons that the overall quality of persons coming into teaching has declined is the decrease in the attraction of highly-qualified females. Other professional opportunities have opened for women. For example, in 1960, 2.5 percent of the law degrees went to women, but 30 percent went in 1980. Women's share of medical degrees rose from 5.5 percent to 23.4 percent in that time.[3]

An obvious implication of this is that low teacher salaries have discouraged many of the most competent graduates from pursuing teaching careers. The most common solution suggested is to raise salaries. This has been done to a modest degree, but teacher salaries cannot now and are never likely to compete with doctors, lawyers, or business person salaries. An alternative to across-the-board higher salaries is merit pay for those teachers demonstrating a high level of competence. The hope behind such schemes is that quality individuals would be attracted to teaching because there would be comparable rewards if

they are truly successful, and they will stay in teaching because the alternatives are not that much more remunerative.

The difficulty is that after several hundred attempts to implement such merit pay plans, only a few remain in existence. The major stumbling block is how to accurately determine teacher excellence. How does one judge how much a teacher has contributed to the education of a classroom of students? Teachers are given varying quality students to begin with. A teacher with a class full of low achievers may actually do wonders with them, yet when the class grades or test scores are compared with other classes, that success does not show. Until our present effort, to be discussed below, this conundrum has negated the attempts to reward superior teachers.

Another attempt at countering the decline in teacher competency has been the moves to increase standards for entry into teaching. Numerous states have moved to require higher grade-point averages for student entrance and exit from schools of education. They have established higher certification standards, standards which in the past have been notoriously low. There are efforts to use a new national test, the National Teacher Examination (NTE), to measure academic preparation for teaching. While such steps raise the academic quality of the students coming into teaching, none of these screens how well a person can actually teach nor do they insure that a sufficient supply of new teachers will be available. If standards are set too high, a severe shortage of teachers may be created. And by not judging teaching quality, the goal of higher achievement by students may not be reached.

This observation leads directly to a confrontation with the issue of exactly what it is that constitutes effective teaching/learning. A 1984 review of the then three main strategies for teacher evaluation found them to be wanting. Pencil and paper tests of teachers, test scores of pupils, or evaluator impressions of teachers all had little ability to accurately assess teacher competency.[4] Efforts in the last five years have attempted to address such deficiencies. The most notable is a "measurement-based approach" that seeks to avoid the subjectivity of teacher rating systems by specifying a set of behaviors thought to be critical for successful teaching and then systematically scoring teachers with respect to these behaviors.[5] A potential flaw in this approach is its circularity. Superior teachers are said to be those who exhibit particular behavior, but those behaviors have not yet been clearly shown to be those which are most responsible for gains in student achievement. Thus, the need for a method of identifying superior teachers and subsequently superior teaching techniques remains.

Yet another theory of why student achievement has been lower relates to the evolution of teachers from professionals to assembly line workers. This argument is that teachers have been trained as professionals to act independently, but as schools have evolved, teachers have been relegated to the role of workers who are expected to produce a product (educated students) using tools and procedures that are prescribed elsewhere. Teachers currently have almost no control over most of the significant aspects of their work, for example, curriculum, instructional periods, class size, materials, tests, and so forth. Such an inconsistency between training and work leads not only to demoralization and burn-out but also low student achievement. The reverse of this, teacher empowerment, is said to be the first step to building teacher competence through self-confidence and independence.

While teacher empowerment may well be an important step, some of its success will depend again on identifying successful teachers and successful techniques. Teachers who are making their own decisions will still need to know more about what works. To date we

have only modest verification of what works and even less application of what works, judging by the overall level of student performance.

Whatever has been done to date to improve the quality of teaching and the level of student achievement is, obviously, not enough. A great deal of disenchantment exists with current public education and with teacher performance in that system. Greater insight into who the better teachers are is critical, as is a better understanding of what it is that better teachers do that makes them successful.

OUR APPROACH

The major stumbling block to identifying superior teachers has been the inability to measure who it is that has really increased student learning the most. Some persons look to the best suburban schools and claim that they have the best teachers because the students there do so well. But are not other factors responsible for doing so well? A great deal of research says that those high levels of achievement are, in fact, related to parents' income, parents' levels of education, parents' occupations, and the educational environment of the home, all of which are likely to be higher in the wealthier suburbs. Also contributing to achievement are factors found to a greater degree in wealthier suburban school districts: higher levels of attendance, kindergarten attendance, going to school with economically advantaged students, and lower pupil-teacher ratios. Given these and other influences, can we really say that the suburban teachers, in this example, are superior?

Obviously, we cannot. Inferring that these suburban teachers are better and that city teachers are worse does not take into account the quality of students coming into the classroom. All that is being measured is the students going out. The real mark of a superior teacher is one who adds value, especially value above and beyond what can reasonably be expected. Suburban teachers have children with higher achievement coming in and, consequently, higher going out. But are these teachers really superior? Have they pushed their students beyond what comparable students learn in the year in which they have them?

Our method of evaluating teachers allows us to take into account the input and the output. We can measure the "value added" to determine whether a teacher, given a pool of students, imparts to them more learning, average levels of learning, or less learning than can be expected, given the progress these and comparable students have made in the past. A superior teacher imparts more learning than can be expected and does so consistently over several years.

The phrase "more learning" is a very relative concept in our method. It is more flexible than a definition that each child should make one year's progress in one year's time because some students could be expected to make less than one year's progress and others, much more. For example, it is entirely possible that a class of gifted and talented students should be expected to make more than one year's progress in a year. The best way to measure the relative amount of learning which has occurred in a given year is to compare how much progress each student makes compared to the progress the student historically might be expected to have made. Thus, a student with lower levels of previous achievement might be expected to have a low level of current achievement. A superior teacher would be one who is able to raise the level of learning beyond that which the achievements of previous years would predict.

In order to evaluate teachers in this fashion our method of teacher evaluation relies on establishing how much progress students have traditionally made each year. Individual

student achievement scores are then compared to their predicted scores. The difference, the so-called residuals, are then aggregated at the classroom level. If the students in a classroom collectively have a positive residual, that is they, on average, learned more than was predicted they would have learned, then we point to the teacher as having been a better-than-average teacher. If the residual is large and positive and shows that students collectively did much better than would be expected, given their individual backgrounds, then we can begin to suspect that we have identified a superior teacher. If we see this same exaggerated level of achievement in a teacher's classroom for several years, our confidence increases that this is, indeed, a superior teacher. By the same token, if a teacher's classroom-wide residual is consistently large and negative, such a teacher may be identified as weak and requiring attention.

This same procedure of comparing individual students with the level of achievement expected of them based on performance can be used at the school level as well to identify superior schools. The only difference is that the calculations are done for all the students in each grade. The predicted performance of each student in a given year is subtracted from the actual performance of each student, and the residuals are then aggregated for the entire grade. If the residual for an entire grade is large and positive, then a superior school at that grade can be identified. Furthermore, if the school consistently across grades produces a large and positive residual, it could be said overall to be a superior school. Note, however, that the school effect measured this way includes the sum of all teacher effects.

The work done to date concentrates on identifying outstanding teachers and outstanding schools in a particular grade. In the future, the test of identifying an outstanding school will be undertaken. What it requires is repeating the work we have done for one grade, the fifth grade, in grades four, three, and two.

In order for the reader to gain a better understanding of exactly how our method of evaluation works, the next section will discuss in detail how we predict how each student should do in a given year based on the performance of preceding, comparable students. This prediction required extensive data collection and complex statistical work. The idea that performance can be predicted is based on some twenty-five years of research by many noted education scholars.

A REVIEW OF STUDENT PERFORMANCE RESEARCH

In 1966 the U.S. Office of Education published a document by Coleman, et. al., that was one of the first to seek to empirically investigate what factors influence student performance. The report is best remembered for its conclusions that " . . . a small positive effect of school integration on the reading and mathematics achievement of Negro pupils after differences in the socio-economic background of the students are accounted for" (pp. 29-30); and "The effect of a student's peers on his own achievement level is more important than any other school influence" (p. 325).

After its publication researchers began questioning the report's conclusions. Coleman's work set off a wave of academic research aimed at explaining the academic achievement of students. A driving force behind the wave was the desire to identify what is responsible for persistent low academic achievement among low socio-economic students. Hundreds of serious studies have been conducted which seek to measure the influence of various factors on student achievement. Some of the findings were cited above. These findings and others cover at least one hundred potential factors.

One way to try to get a handle on the numerous potential factors is to group them into categories as some researchers have.[6] Among the more encompassing categories of influence are: (1) individual student characteristics such as age, sex, race, pre-school experience, and school attendance; (2) family characteristics such as family size, parents' education and occupational status, family income, and emphasis on education in the home; (3) peer group inputs such as social class composition, racial composition, ability composition, and classmate or school annual rates of turnover; (4) teacher input, such as amount and type of education, recency of education, certification and tenure, experience, verbal achievement, and so forth; (5) school characteristics such as expenditure per pupil, teacher salaries, class size, teacher turnover, age and condition of the school building, and so forth; and (6) past student achievement, such as scores on national tests.

Rather than describe in detail what the research has shown to be the influence of each of these factors, their impacts are summarized in Table 1 as to both whether a relationship exists and the direction of that relationship.

TABLE 1

VARIABLES FOUND TO BE RELATED TO STUDENT ACHIEVEMENT

Individual Student Characteristics	Direction of Relationship	
Sex (Female)	Reading	+
	Math	-
Preschool Experience		+
Age (Above Classroom Average)		-
Attendance		+
Effective Variables (self-concept)		+
Family Background		
Family Size (Large)		-
Parents' Education		+
Parents' Occupational Status		+
Family Income		+
Parents' Expectations		+
Peer Group Input		
Social Class Composition (High)		+
Racial Composition		Mixed
Ability Composition (High)		Mixed
Turnover in Classroom		Unclear
Teacher Input		
Education		Unclear
Recency of Education		Unclear
Experience		+
Specialization		+
Verbal Ability		+
Sex (Men and Low-Income Students)		+
Race (Black Teacher - Black Students)		+
(Black Teacher - White Students)		-
Schools		
Expenditure Per Pupil		+
Ability Grouping		Unclear

(Table 1 Continued)

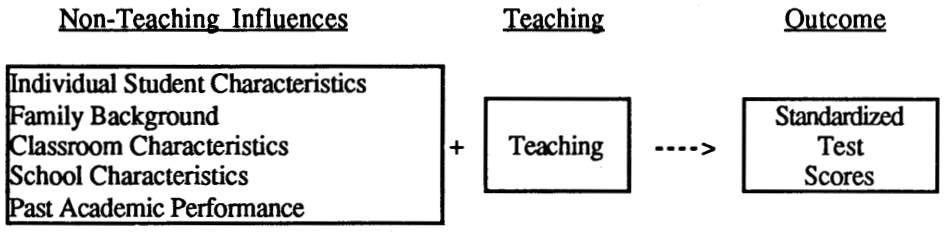
Guidance Services	+
Teacher Turnover	-
Teachers' Salaries	+
Past Performance	
Past Achievement	+

Thus, for example, research generally shows being female, having preschool experience, attending school regularly, having a good self-concept, having well-educated parents, having parents with high occupational status, high income and high expectations for student performance all contribute to student achievement in a positive way. Also contributing in a positive fashion are: a peer group of high socio-economic background; teachers with more experience, high verbal ability, and a specialization; attending schools which spend more than average per pupil and which have guidance services, pay teachers more, and have longer school years. Another key to student performance, according to preceding research, is past achievement by the individual student.

Research has also shown some variables, such as being above the classroom average in age and coming from a large family, to be negatively related to student achievement. For some other factors, such as rate of student turnover in a classroom, the recency of a teacher's education, or the size of the physical plant of a school, the research reveals unclear results. That is, the studies to date cannot show definitively what impact these variables have on achievement.

The problem with the research to date, however, is not one of identifying variables which explain student academic performance, but rather determining which variable has the greatest explanatory power. The limited number of variables examined in any one of the previous studies has made such a determination impossible. But the extent of the literature and the success others have had in identifying factors which are influential suggests that the project we propose is feasible.

If we can identify most of the non-teacher-related determinants of student achievement, we might be able to assume that some portion of achievement not determined by demographic, environmental, and other such characteristics is determined by what goes on in the classroom -- mainly teaching. A model illustrating this hypothesis is shown below.



Essentially what we are suggesting is that the measurement of the influence of teaching is a residual. If non-teaching influences explain 80 percent of the variance in student performance, then teaching might explain some portion of the unexplained variance. This is our assumption.

At first glance it seems impossible to attribute unexplained variance in student achievement to teaching. After all, this assumes that we have identified all of the pertinent non-teaching variables that explain student performance so that the residual could only be the effects of teaching (and noise). And what if Johnny has a cold on the day of the test? Or what if Mary is at her peak on the day of the test? We can only assume that these are

random occurrences and in a large data set will not influence our ability to identify systematic relationships.

We can never hope to prove beyond the shadow of a doubt that the residual is due only to teaching, but it may be that we can present evidence so compelling that the logical conclusion is that the residual effect is in large part teaching.

Let us pose a hypothetical case. Assume that we are able to identify and measure non-teaching variables which explain 75 percent of the variance in student performance on an academic achievement test. This leaves 25 percent of the variance unexplained. What if we retrospectively use the non-teaching variables to predict how students will do on this test and compare this to actual performance (generate a residual)? What if we aggregate these residuals by classroom and find that Ms. Elliot's fifth grade class, on average, scored five percentage points higher than we predicted? And what if we find that every one of Ms. Elliot's classes have, on average, scored five to ten percentage points better than predicted every year over the past five years? And what if we found that Ms. Elliot had moved from a low SES school to a high SES school two years ago? Doesn't logic tell us that Ms. Elliot's consistently superior results must be based on her outstanding teaching? That is the proposition we will test.

THE RESEARCH PLAN

In order to begin, we first had to develop a means of explaining a large portion of the variance in student achievement. To simplify this task, we used only elementary schools. In such schools one teacher is basically responsible for the academic achievement of his/her students. We qualify the term responsible with basically because there are always instances where children having special difficulties (e.g. reading) receive additional assistance from a specialist and because subjects such as art, music, and physical education may be taught by other teachers. But one teacher is with the students the majority of the time.

We further focused our initial efforts by concentrating on predicting fifth grade achievement. Fifth grade is the end of elementary school for many children, and it is the year in which national tests are given to many students, including those in our study district. Our study district is an urban district with a mix of races, a history of lower student achievement, especially among minorities, and the usual liturgy of conditions found in cities and city schools. Fifth grade is also chosen because of the ability to put together several years of history on each student. With more data, it was thought, our ability to predict achievement would be increased.

The basic measure of achievement chosen for this experiment was reading level as revealed on national, standardized tests. While many persons are critical of such tests, especially because of generally lower scores for minority children, national tests provide the only consistent measure of achievement currently available. Our study district also employed a reading series which required teachers to regularly evaluate reading achievement. Since national exams are not given annually, we used these "semi-standardized" evaluations for measuring achievement in the school years between national tests and to "validate" their use by comparing such scores with national achievement test scores. We found, unfortunately, that the scores on the semi-standardized tests only partially correlated (.64) with the national test scores at fifth grade. A better measure might be annual, content-based learning tests that would test material actually covered in the school year. But such exams are not currently available. The thrust of our effort at this juncture was to test the basic concept of predicting student performance, so we used the best available indicator at the fifth grade level, the national test reading score. The lack of a

high correlation between the national tests and annual reading tests did not affect our results at fifth grade but other, better measures might be employed in future efforts.

The research plan called for a retrospective examination of students (and teachers and schools) who had taken the national, standardized tests at the end of fifth grade in (1) 1986 and (2) 1987. We confined ourselves to two years because of scale (over the six year period of kindergarten through fifth grade over 20,000 students were enrolled at some point in these two classes) and because our study district only had used the reading series and "semi-standardized" tests since 1983. We had available to us, therefore, the immediate past reading achievement data for two years to help predict the performance of the 1985-86 fifth graders and three years of scores to help predict the performance of the 1986-87 fifth graders. Because central records were not kept of elementary school grades, we could not use grades or grade-point-averages as predictor variables. The reading series yielded the only generally available measure of annual achievement.

INDEPENDENT VARIABLES

The reading test scores were used as independent (possibly predictor) variables for the fifth grade. When analyzing the fourth and third grades, they became the dependent variable, the variable we sought to explain. Our identification of potential independent variables yielded a list of over 100 factors, such as individual, family, classroom, and school characteristics, as well as past academic achievement. The number of independent variables was obviously reduced in each preceding year because one cannot use a fourth grade reading score to predict a third grade reading score. The independent variables and projected uses for each grade are shown in Table 2. This is a condensed list, but it covers the major variables examined.

TABLE 2

INDEPENDENT VARIABLES USED TO PREDICT STUDENT PERFORMANCE

Independent Variables	Year Applies				
	Grade 5	Grade 4	Grade 3	Grade 2	Grade 1
Individual Characteristics					
Gender (student is male)	X	X	X	X	X
Race					
White (student is White)	X	X	X	X	X
Black (student is Black)	X	X	X	X	X
Hispanic (student is Hispanic)	X	X	X	X	X
Exceptional Education (percent of years student has had exceptional education flag)	X	X	X	X	X
Age (relative age of the student compared to classroom peers)	X	X	X	X	X
Student Attended Kindergarten	X	X	X	X	X
Student Attended Private School	X	X	X	X	X
Previously Attended Specialty School	X	X	X	X	X
Attends Specialty School	X	X	X	X	X
Family Characteristics					
Single Parent (student lived with single parent)	X				
Poverty					
AFDC and/or Food Stamps (current year)	X	X	X	X	X
Cumulative Scale of AFDC and/or Food Stamps (pct.)	X	X	X	X	

(Table 2 Continued)

Eligible for Free or Reduced Lunch (current year)	X	X	X	X	X
Cumulative Scale of Free or Reduced Lunch (pct.)	X	X	X	X	X
Mobility					
(Total number of schools student attended per yr.)	X	X	X	X	X
(Cumulative number of schools attended av. per yr.)	X	X	X	X	X
Neighborhood					
(Percent poverty in home census tract)	X	X	X	X	X
Percentage of Males in Class	X	X	X	X	X
Percentage of White Students in Class	X	X	X	X	X
Percentage of Black Students in Class	X	X	X	X	X
Percentage of Hispanic Students in Class	X	X	X	X	X
Percentage of Same Race Students in Class	X	X	X	X	X
Average Rank of Peers on Second Grade Iowa Test:					
Total Reading Grade Equivalent	X	X	X		
Total Math Grade Equivalent	X	X	X		
Total Reading Percentile Rank	X	X	X		
Total Math Percentile Rank	X	X	X		
Average Scott-Foresman Reading Level of Peers	X	X	X	X	X
Percentage of Exceptional Ed Flags Among Peers	X	X	X	X	X
Percentage of Peers Whose Parents Receive AFDC	X	X	X	X	X
Percentage of Peers Whose Parents Receive Food Stamps	X	X	X	X	X
Percentage of Peers Who Are Eligible for Free Lunch or Reduced Lunch	X	X	X	X	X
Percentage of Peers Who Have Been Held Back	X	X	X	X	X
Average Classroom Attendance	X	X	X	X	X
Average Number of Schools Attended	X	X	X	X	X
School Characteristics					
Pupil Percent Stability	X	X	X	X	X
Pupil Percent Minority	X	X	X	X	X
Pupil Percent Attendance	X	X	X	X	X
Pupil Percent Above Average in Grade	X	X	X	X	X
Teacher Percent Minority	X	X	X	X	X
Percent Teachers Holding Masters Degree or Beyond	X	X	X	X	X
Percent Teachers In Service Less Than 2 Years	X	X	X	X	X
Percent Teachers In Service Less Than 2-5 Years	X	X	X	X	X
Percent Teachers In Service 6 or More Years	X	X	X	X	X
Academic Performance					
Fourth Grade Scott-Foresman Reading Level	X				
Third Grade Scott-Foresman Reading Level	X	X			
Second Grade Scott-Foresman Reading Level	X	X	X		
First Grade Scott-Foresman Reading Level	X	X	X	X	
Second Grade Iowa Tests:					
Total Reading Grade Equivalent	X	X	X		
Total Math Grade Equivalent	X	X	X		
Total Reading Percentile Rank	X	X	X		
Total Math Percentile Rank	X	X	X		
Kindergarten Pre-Reading Tests 1-6 (sum of)	X	X	X	X	X
Kindergarten Pre-Reading Tests 5-6 (sum of)	X	X	X	X	X
Attendance	X	X	X	X	X
Number of Years Held Back	X	X	X	X	X

METHODOLOGY

A number of statistical techniques were used to sort and analyze the data. But the key statistical tool used to explain the variance (the difference) in achievement test (or the

alternative semi-standardized test) scores was multiple regression. This is a technique which both identifies those variables that have an influence on the dependent variable (test scores) and then rank-orders them with a numerical declaration of how much impact each has in explaining the test scores. We entered the independent variables into multiple regression equations to explain the variance in test scores for year one (fifth grade in 1985-86), year two (fifth grade in 1986-87), and both years combined. We then took the factors identified as influential and used them to predict individual student performance. It is these predictions which were then compared with actual performance on a student-by-student basis. The differences between the two, the residuals, were then aggregated on a classroom and school basis to identify specific teachers and schools.

THE PROBLEM OF MISSING DATA

While our study district was able to provide us with rather extensive data, the student records were not perfect. Most students, in fact, did not have complete data in their records, most often because they were not in that school district for the entire six year period but sometimes because pertinent data were not recorded in their file either by oversight or its not having been provided. This forced us to develop statistical techniques for filling in the data gaps in a fashion which would do the least damage to validity. The results were of more than sufficient quality to test the method, but we would strongly state that any efforts to apply our evaluation method on an actual as opposed to an experimental basis should have a more complete data set than we had. The results would be more defensible.

CLARIFICATIONS OF METHOD

Before describing how our predictive equation is constructed and how it, in turn, is used to predict student scores, several possible misperceptions deserve clarification. First, it is important to stress that we do not claim to be testing any particular model of causation. We merely search for variables which are independent of the current year inputs of school and teacher that correlate with test performance. Second, in the absence of a strong theoretical model of causation, the particular equation derived for our study district in the two years in question may not be the most appropriate in other school districts or years. In addition, other districts may have access to different variables, making it impossible to replicate our predicting equation, even if it were the appropriate one. But our proposed approach is a flexible one which can be customized in each school district rather than one in which all districts merely use the results of our initial study.

Third, we emphasize that this method provides a ranking of teachers and schools relative to one another, not an absolute ranking relative to a standard of excellence. We can identify a ranking of schools and teachers based only on relative student performance which, in turn, is based on a history of performance of comparable students within the study district. This does not preclude the establishment of higher standards of achievement and the ranking relative to those higher standards. But the new standards will be expressed in terms of achievement above and beyond that predicted, not in relation to any outside norm.

Fourth, our initial step of identifying superior teachers does not include an examination of what makes them successful. That requires a whole other study of those identified as outstanding. Such a study should certainly follow successful application of our method. In the meantime our research is motivated by past findings which strongly demonstrate that teachers and schools do differ greatly in their effectiveness.

THE PROCESS: THREE STAGES

The process of ranking schools and eventually teachers involves three stages. In the first we search for the estimating equation (the mix of variables) which best explains the variance in test scores across the students for each grade, starting with the fifth grade. The second stage involves the application of the estimating equation to each student to determine her/his expected performance in each year and the subtraction of this expected performance from her/his actual performance, creating the residual or difference between actual and expected performance. The third stage involves aggregating those individual residuals across classrooms to compute teacher impact and across grades and schools to compute teacher and school impact by grade and by school. These residuals can then be ranked from the most positive to the most negative and can be the basis for further analysis of the statistical significance among teachers and among schools.

STAGE ONE: CREATING THE ESTIMATING EQUATION

To begin the first stage we identified three broad sets of controls upon which to condition the school and teacher-specific effect. First, there are student characteristics, such as race, gender, and socio-economic status, that are beyond the influence of schooling and which need to be taken into account. Second, to prevent the school and teacher from receiving credit for the natural ability of the students they happen to inherit, we use a broad set of previous test results which predate the influence of the current year of schooling. Third, because "peer" variables, such as the percent of low income or low achievers in one's school or grade, have been suggested as important influences on achievement, we have generated "peer" variables which capture the current composition of the school and grade.

To these three we have added one additional set of variables, a set designed to overcome the data limitations caused by missing data. We created a scheme whereby mean values were assigned to missing data, so that we could use a larger number of student records without jeopardizing the validity of the research effort. And we could analyze whether those with missing data have systematically different performances.

Having identified these controls and having created with original data and newly generated data (i.e. "peer" variables, missing data means, and dummy variables) almost 500 variables that might explain the variance in student achievement, we wanted to run a stepwise regression on all variables. Such a technique would identify every variable which correlates significantly with student performance and would directly provide the predictive equation which explains the largest share of the variance in student performance. Unfortunately, our computer resources could not handle such a massive task, so we had to divide the problem into sequential sub-problems. We ran regressions on subsets of the data to locate the most important variables in each subset. These stronger variables were then run together to identify the strongest among them. We cannot be certain, but we think that our subsequent approach did not cause much variation in the final predictive equation.

What emerged from our regression was a listing of factors that are related to student achievement. Several of the more important variables are listed in Table 3. The linear equation used provides coefficients whose magnitude can be easily understood. To facilitate understanding, each of the examples will be discussed. Before doing so, however, we must state that we do not pretend to know what really causes students to score well or poorly. For example, the seventh item on our list, whether a student is black, has an apparent negative impact on test scores. That does not indicate that black persons will always score 5.67 points lower. It just states that there is a relationship here between

race and score; it may have everything to do with income rather than skin color. We are not able to discern.

At any event we will explain what the more important variables are that relate to fifth grade reading test scores. The variables which are the strongest predictors of how a student will do on the fifth grade Iowa Test of reading skills are listed in Table 3. Two statistics are listed, a beta and an unstandardized regression coefficient. The beta is a measure of how strong a predictor

TABLE 3

EXAMPLES OF THE BEST PREDICTORS OF STUDENT ACHIEVEMENT

Variable	beta	Unstandardized Regression Coefficient
1. Reading test score, Grade 4	25.84	10.45
2. Age as of Sept. 1 in year student attended 5th grade	17.09	-7.40
3. Average score of other 5th grade students at school on 5th grade Iowa Test - Reading - Grade equivalent	13.15	1.24
4. Average score of other 5th grade students at school on 5th grade Iowa Test - Math - Percentile rank	11.52	-1.60
5. Average score of other 5th grade students at school on 5th grade Iowa Test - Math - Grade equivalent	11.32	3.94
6. Sum of pre-reading tests 5 and 6	10.06	1.16
7. Whether the student is Black	9.98	-5.67

each factor is in comparison to all other factors. For example, a beta of 20 is twice as powerful as a 10. The unstandardized regression coefficient states the dimensions of the relationship between a change in the factor and how such a change influences the test scores.

As is clear from the listing, the factor that is the best predictor of the fifth grade reading score on the Iowa test is how well a student did on reading in grade four, as measured by their assessment on the Scott Foresman reading series. The regression coefficient shows that for every year above or below the expected level of reading at the end of grade four, which in this case is 5.0, the start of fifth grade, a student's Iowa Test score should be 10.45 points higher or lower, on average. Thus, students who at the end of grade four are assessed to be reading at level 6.0 should, on average, have an Iowa Test score (percentile rank) which is 10.45 points higher than students who were rated a 5.0 reading level, other variables held constant. And students who were rated at 4.0 reading level should have Iowa Test scores 10.45 points lower, on average, than those who were ranked at 5.0, other variables held constant.

The next best predictor is the age of the student when he or she entered the fifth grade. The older they are, the worse they are expected to do. The coefficient indicates that for each year older than normal for the grade, a student's score is expected to be 7.4 points lower, other variables held constant.

The third best predictor is the average score of the other fifth graders on the Iowa Test of reading. For each grade equivalent higher than fifth grade the average lies, a student's score should increase 1.24 points. This is a strong but modest relationship.

The fourth and fifth strongest predictors are related and a bit puzzling. They are each ways of measuring the average score of all other fifth graders on the math portion of the fifth grade Iowa Test. The percentile rank is negatively related but the grade equivalent is positively related. The grade equivalent has a greater impact and indicates that for each grade above fifth that the average score lies at a school, a student's reading score is expected to be 3.94 points higher.

The sixth factor on the list is the sum of two test scores from the kindergarten pre-reading test. The regression coefficient indicates that for each step in readiness, the reading score is expected to rise 1.16 points.

The seventh factor, being black, has already been discussed. The findings are that a black student, for whatever the reasons, is expected to score 5.67 points lower than a white student, all other factors held constant.

As the list continues, several broad patterns also emerged. First, of the forty-five variables in the linear equation which show up as significantly related, thirteen are previous test scores, eight are measures of age in grade or attendance, seven are measures of socio-economic status, and five are ethnicity or race. Second, a large share of the variance, a serviceable 62 percent, is explained by the variables we have included in our study. Our methodology posits the remaining 38 percent to school and/or teacher effects (and random factors). Third, our findings that specialty school students perform better, students with low socioeconomic status perform worse, and so on, yield a general tenor of results that fits with intuition and many previous results.

We should stress that the predictive equation produced is specific to our study district and may benefit from the introduction of additional variables. Yet, such a predictive equation could be reproduced for other school districts, taking into account variables thought to be particularly important to those districts. The approach we describe is flexible enough to be modified for other districts.

THE SECOND STAGE: CREATING RESIDUALS

Having created a predictive equation, we then have to apply it to each student's record. This means that the data required by the estimating equation is inserted, and the equation is solved to yield an estimate of the score each student is expected to receive at the end of fifth grade. These estimates are then subtracted from the actual scores achieved on the tests given at the end of the fifth grade to create the residuals. These residuals form the basis for examining which schools and, ultimately, which teachers have been unusually effective.

THE THIRD STAGE: EXAMINING RESIDUALS

In order to rank schools from those that added the most unexpected value to those that delivered much less than expected, we aggregate the individual student residuals by grade and school. Schools with large positive residuals have students who were predicted to do worse than they actually did in fifth grade. Alternatively, schools with large negative residuals have students who were predicted to do better in the fifth grade than they actually did. Thus, the difference between predicted and actual performance is the measure of the school's contribution at the fifth grade.

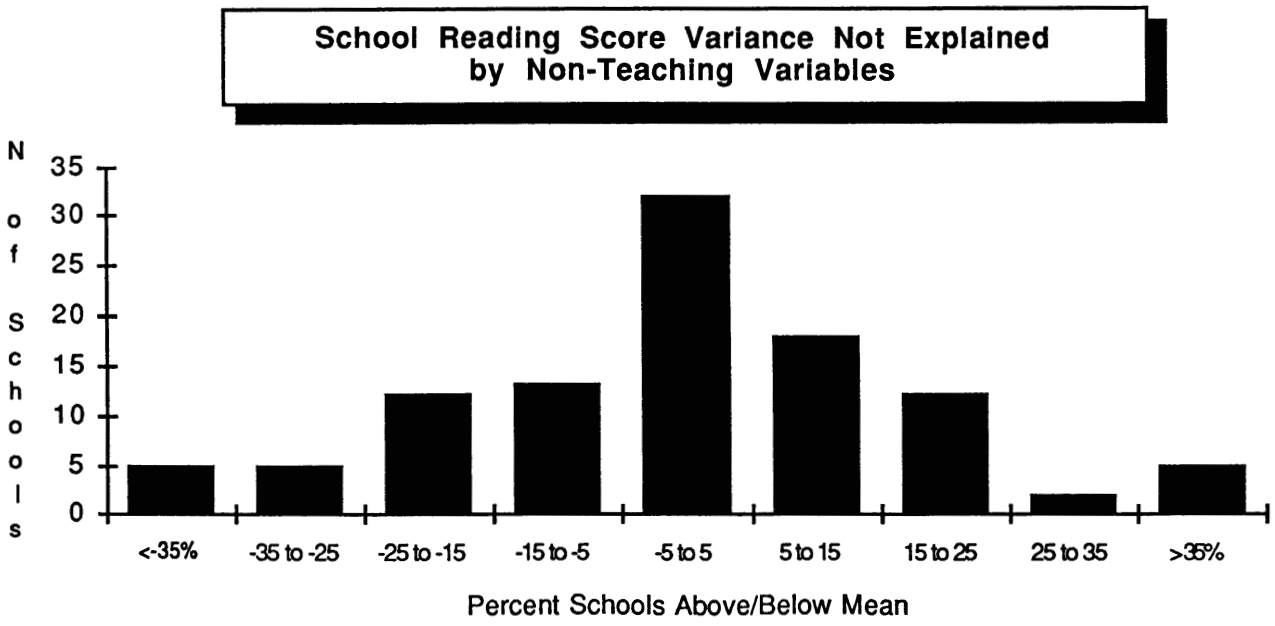
A negative residual does not indicate that a school (and its fifth grade teachers) did not add value. The negative residual indicates only that the school did not add as much value

as expected, given the characteristics of the students and how they performed in previous years. A negative fifth grade residual may be due to either of two factors: a weaker contribution by a school's fifth grade teachers or a stronger record of achievement before fifth grade with a tailing off in performance at fifth grade. Regardless, the method does allow us to rank schools based on their relative contributions to fifth grade performance. Subsequently, the same procedure can be followed to rank the lower grades and to then aggregate all grades to identify the schools which in a sum of all grades have a net positive residual or which have a large net positive residual for each year. These would be truly outstanding schools.

We must note that at this juncture we do not distinguish between the teacher and school effect. The school effect remains a composite of the influence of the particular teachers in the school, all other influences of the school, and other factors not captured by the original set of variables. When we focus on teachers below, we modify this procedure to take out the school effects.

The generation of the school-wide fifth grade residuals yielded a distribution of schools with scores which ranged from a negative 8.1 at School M to a positive 9.5 at School C. As can be seen in Figure 1, the majority are centered around zero, showing they performed about as expected. But there are extremes. To discover which schools have residuals significantly greater or lesser than zero, we set some extreme limits which identified those schools which are at least 25 percent better or worse than average. Out of the almost one hundred schools in the sample, seven schools are identified as having unusually large, positive residuals, and ten are identified as having unusually large, negative residuals. These ten schools performed markedly worse than predicted, and the seven performed substantially better. Certainly if another standard were used, the number of schools identified would enlarge or diminish. The main conclusion is that this method does yield a ranking of schools by the unexpected value they contribute. Both the best and the worst schools can be identified and recognized for their contributions (or lack thereof).

FIGURE 1



EVALUATING TEACHERS

The previous section describes a flexible framework with which to identify specific school effects on the performance of students. To evaluate teachers we must isolate the influence of individual teachers (classrooms) from the broader influence of the schools in which those classrooms are located. We need to break down the "composite" effect of both classroom and schools that was used to evaluate school influence. This is critical for any effort to ultimately tie student performance to specific teacher inputs.

We must note that basically what we are doing at this point is measuring the "classroom" input, not just the teacher input. We call it the teacher input because that is largely, if not exclusively, what it is. But in most school systems today, even in elementary schools, the primary teacher may not be with the students for the entire day. Special subjects taught to each class may require other teachers or a sharing of responsibility. And "classrooms" may be subdivided for specific exercises or learning experiences. Thus, we may not fully measure the influence of individual teachers. But the primary teachers are most involved in the skills most directly tested, and they do help to orchestrate the other teachers. The fact that we may not measure teacher impact precisely demonstrates that our results are limited by our data not by our framework.

In order to separate the school influence from the teacher influence, we had to recalculate our estimation equation. The major change was the inclusion of dummy variables for all of the over 100 schools in our sample to separate out the impacts of the schools. We then ran a stepwise regression as we had before on all of the students and schools in our data set. In this regression, 50 of the school dummies emerged as statistically significant with an r-squared of .130. The schools did have an impact that should be noted.

To generate our final estimating equation, we did include these significant school dummies along with those variables which had previously been used in the final estimating equation for the schools. While all students were included in the original estimation, our final equation involves only the 6249 students for whom we could identify a classroom. These students constitute about two-thirds of our original sample. Those variables which were selected as significant became the new estimating equation for classroom impacts. The school dummy variables hold constant the school effects, yielding an ability to examine the effects of the individual teachers (classrooms).

Before describing the results, we should make two additional, modest caveats. First, we limit ourselves to examining only the fifth grade reading scores, although other test scores could be examined in the same fashion. Second, we know our classroom data are for the first year a student is in his or her fifth grade. Since approximately four percent of our sample has repeated the fifth grade, we have a very modest misallocation of data. This should not substantially affect our results.

RESULTS

The final estimating equations for the fifth grade reading scores do not differ substantially from those estimated for the schools. Only six of the school dummies emerge as significant, and only nine variables are excluded from the earlier estimation. The explanatory power remains roughly the same with an r-squared of .614. And most of the variables included in both equations have very similar coefficients.

But another wrinkle does influence our procedure. While we have data on over 6000 students by classroom, we cannot use all of those observations. Many of them are in

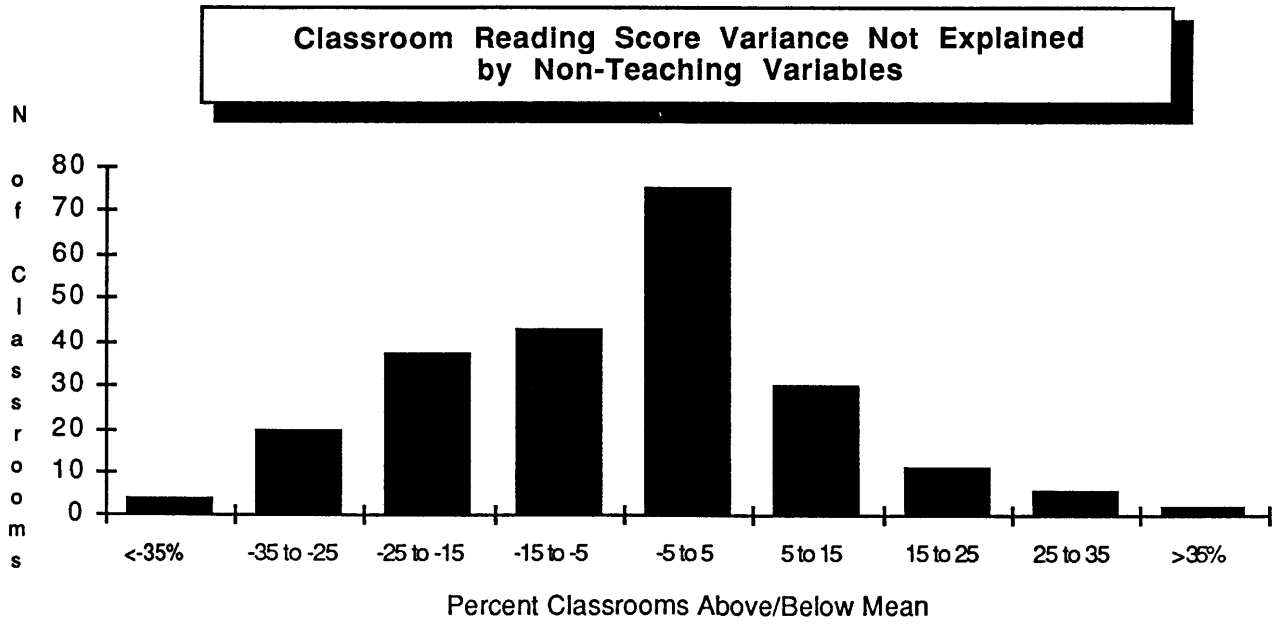
classrooms with too few observations to draw reasonable inferences. For purposes of our examination, we use data on only those classrooms which have at least twelve observations. This is an arbitrary number, but it would clearly be unreliable to base our examination on classrooms with just a few observations. They could not reasonably be assumed to be representative of the entire class of 25 to 30 students. (This was not a problem at the school level because the minimum number of cases was 31 and the mean was 103.)

The estimating equation was used to produce predictions for each student. These, in turn, were subtracted from the actual test results to yield residuals. The residuals were then aggregated by classroom within each school. To identify those teachers who were outstanding in either a positive or a negative sense, we isolated those classrooms in which the residual was more than twenty-five percent away from the mean of approximately zero.

A total of 228 classrooms emerged with 12 or more observations. As is shown in Figure 2, to an even greater degree than at the school level, classrooms generally cluster near zero, indicating performance very close to that predicted. But 32 were identified as outliers, with 8 positive and 24 negative (Figure 2). These are the classrooms whose teachers we would begin to focus on as outstanding in each direction. Interestingly, two classrooms in one school were identified as outliers, one positive and one negative.

To be more sure that a teacher is, in fact, an outstanding teacher, it would be much safer to use multiple years of analysis. This would help to reduce any statistically aberrant results. We might define an outstanding teacher as one who has students with substantial positive results at least three years in a row. It may be possible to reduce the need for consistent performance if, through use, we can show that large year-to-year variations in residual per classroom are uncommon. Initially, though, multiple years of success would be the best measure.

FIGURE 2



Having proven that this method can identify teachers that add unexpected value at grade five, we need additionally to prove it can work in grades one through four, and in grades six through eight where there is still one primary teacher per classroom. The basic approach would be the same. The only difference for the lower years is that there would be less data available as we work down from grade five. This may significantly reduce the power of the predictive equation. We will not know by how much until we undertake the exercise.

It may also be possible to use a variation in this technique to evaluate middle and high school teachers who teach in specific subject areas. What would be needed are two elements. We would need to develop specific measures of annual progress. This may come from content-based tests of the material in a subject area or in some cases the results of Advanced Placement exams. The second element needing development is a history of student progress so that a predictive equation can be developed based on what a comparable group of peers has been able to accomplish.

Another potential variation is a different set of standards for identifying outstanding teachers in any of these efforts. We have used a definition for both teachers and schools of choosing those whose residuals are at least two standard deviations from the norm. That is an extreme. Other possible methods include identifying the best 100 teachers in the district or labeling as outstanding those teachers more than one or one-and-one-half standard deviations above the norm. Our approach remains flexible and can accommodate a range of variations.

CONSISTENCY OF RESULTS

How confident are we in the results? If there are indeed differences between good and poor teachers and thus good and poor schools, we would expect the results to be consistent across the years. That is, if a teacher is found to be very poor in one year, we can logically expect that teacher to also do poorly the following year. (Although given the fact that our data are really for classrooms, the relationships would not be perfect. Teachers change rooms or grades, move, retire, are ill for extended periods of time, etc.) To test this consistency we initially attempted to correlate the mean residual scores by classroom for the two years. Unfortunately there were only 27 cases where we had records on 12 or more students in the same classroom for both years -- a number much too small to analyze.

We were, however, able to correlate the results by school for 103 of the schools. The resulting correlation coefficient was .581, indicating a good deal of consistency by school from one year to the next. Furthermore, we examined a scatter diagram of the residuals and visually confirmed the consistency. No schools with really poor residuals in the first year dramatically improved in the second year or vice versa. We found the same to be true concerning the high positive residuals.

CONCLUSIONS

The most important conclusion is that teachers can be evaluated using this method of predicting student performance and comparing it with actual outcomes. Our experiment showed that we can differentiate among teachers on the basis of how much value they have

added in comparisons with what comparable students ordinarily learn. We can identify the best and the worst teachers. Schools can also be identified by how much each grade of teachers contributes and how much the school in total contributes beyond the norm for a similar mix of students. The application of this method would substantially increase the accountability of individual teachers as well as schools.

We must admit, however, that in order to be applied in a real as opposed to an experimental situation the quality and completeness of the data on each student would have to be substantially improved. With large amounts of missing data, our equations were able to account for approximately 62 percent of the variance in student achievement scores. In social science research such a figure is considered meaningful. But with better data, that figure could likely be closer to 75 percent and carry even greater credibility. Also requiring improvement is the measure, national test data. Local, content-based learning tests might prove to be superior, but they are not yet sufficiently in place to be judged.

The procedure also needs to be used at both lower and higher grades to determine how good the predictive equations are. It is likely that at each lower grade with one year's less information, the power to predict will diminish. But without experimenting with it, we will not know. In the near future, such experimentation will occur, and we will know how valid this method is using existing data.

We should also note that identifying the best and worst teachers and schools is only the first step in the struggle to improve educational quality. Having identified the best teachers, at least three subsequent steps should be taken. First, these teachers should be monitored to determine what it is that they do which makes them so successful. It is precisely this information which should get passed on to others as quickly as possible. Second, those that desire it should be elevated to master teacher status so that they can directly convey their methods of teaching success to others. Third, the successful teachers should be financially rewarded for achievement. Past efforts to reward merit have foundered because those most deserving of it could not be identified. Our method has broken this barrier, takes into account the quality of the students coming into each classroom, and identifies those teachers who really get students to learn more than similar kids, on average, learn.

At the other end of the spectrum, for those teachers who are shown to contribute much less than expected, a range of options is open. A district could ignore such findings (although at some peril if the method is formally adopted), use the findings to identify and counsel such teachers to improve their teaching, or identify and work to terminate them. Which options are elected is open to local discretion.

That local discretion will certainly be influenced by local teachers and teachers' unions. There may be some union opposition to the adoption of our method because it explicitly attempts to differentiate among teachers. But in other communities such as Rochester, N.Y., that is perfectly acceptable to the union. And some national leaders like Albert Shanker have endorsed the concept of making such distinctions. Adoptions of our method will not be an easy sell. But with growing pressure for accountability before more resources will be given to education, adoption of this method allows both teachers and students to benefit.

FOOTNOTES

1. Gordon S. Black, "The Lack of Confidence in Public Education in Wisconsin (Milwaukee: Wisconsin Policy Research Institute, April 1989).

2. Testimony before the Governor's Commission on Secondary Schools in Florida as quoted in B. Frank Brown, Crisis in Secondary Education: Rebuilding America's High Schools (Englewood Cliffs, NJ: Prentice Hall, 1984).

3. Emily Feistritzer, The American Teacher (Washington, DC: Feistritzer Publications, 1983).

4. Glen E. Robinson, Paying Teachers for Performance and Productivity: Learning from Experience, ERS Concerns in Education Series (Arlington, VA: Educational Research Service, May, 1983); and "Incentive Pay for Teachers: An Analysis of Approaches" (Arlington, VA: Educational Research Service, March 1989).

5. Donald Medley, Homer Coker, and Robert Soar, Measurement-Based Evaluation of Teacher Performance: An Empirical Approach (New York: Longman, 1984).

6. R. Gary Bridge, Charles M. Judd, Peter R. Moock, The Determinants of Educational Outcomes: The Impact of Families, Peers, Teachers, and Schools (New York: Ballinger, 1979).

ABOUT THE INSTITUTE

The Wisconsin Policy Research Institute is a not-for-profit institute established to study public policy issues affecting the state of Wisconsin.

Under the new federalism, government policy increasingly is made at the state and local level. These public policy decisions affect the lives of every citizen in the state of Wisconsin. Our goal is to provide nonpartisan research on key issues that affect citizens living in Wisconsin so that their elected representatives are able to make informed decisions to improve the quality of life and future of the State.

Our major priority is to improve the accountability of Wisconsin's government. State and local government must be responsive to the citizens of Wisconsin in terms of the programs they devise and the tax money they spend. Accountability should be made available in every major area to which Wisconsin devotes the public's funds.

The agenda for the Institute's activities will direct attention and resources to study the following issues: education; welfare and social services; criminal justice; taxes and spending; and economic development.

We believe that the views of the citizens of Wisconsin should guide the decisions of government officials. To help accomplish this, we will conduct semi-annual public opinion polls that are structured to enable the citizens of Wisconsin to inform government officials about how they view major statewide issues. These polls will be disseminated through the media and be made available to the general public and to the legislative and executive branches of State government. It is essential that elected officials remember that all the programs established and all the money spent comes from the citizens of the State of Wisconsin and is made available through their taxes. Public policy should reflect the real needs and concerns of all the citizens of Wisconsin and not those of specific special interest groups.