8: TEACHER CHARACTERISTICS

EXECUTIVE SUMMARY

RESEARCH FINDINGS

Traditional psychometric techniques (using ability, achievement, other paper-and-pencil tests, GPAs, and the like) to predict teaching effectiveness (in terms of student achievement) have failed. Certification status appears to be causally related to improved student achievement: regularly certified teachers produce higher student achievement than non-certified or emergency certified teachers. Teacher experience generally has been shown to be positively related to student achievement when other variables are statistically controlled. Little research has been published on the unique characteristics of teachers that make them successful in teaching children in poverty.

RECOMMENDATIONS

- Paper-and-pencil tests are not useful predictors of teaching candidates' potential to teach successfully and accordingly should not be used for that purpose.
- A teaching candidate's academic record (e.g., GPA) is not a useful predictor of his or her eventual success as a teacher. A candidate's record of success in pre-service (undergraduate) technical courses (mathematics and science, for example) may contain useful information about that candidate's success in teaching secondary school mathematics and science.
- Other things equal, 1) students of regularly licensed teachers achieve at higher levels than students of emergency certified teachers; and 2) more experienced teachers produce higher student achievement than less experienced teachers. Teacher selection policies should reflect these facts.
- The selection of teachers who will best contribute to their students' academic achievement should focus on peer and supervisor evaluation of interns, student teachers, substitute teachers and teachers during their probationary period.

8: TEACHER CHARACTERISTICS

BY GENE V GLASS
ARIZONA STATE UNIVERSITY

INTRODUCTION

How can one identify in advance of a decision to hire which teachers will most improve their students' measured achievement? What are the characteristics of promising teachers that will permit an accurate prediction of their ability to teach children well?

This review deals with those characteristics of teachers that might be identified and used in the initial hiring of teachers to increase their students' achievement. These characteristics can include qualities of teachers that are viewed as personal – such as mental ability, age, ethnicity, gender and the like – or as "experiential" – such as certification status, educational background, previous teaching experience and the like. Some characteristics are combinations – in unknown amounts – of personal and experiential qualities, e.g., candidates' performance on teacher-certification tests such as the National Teacher Examinations and state-mandated tests.

This review will not examine characteristics of teachers that would be impractical to assess in the initial hiring and selection process, such as deep personality traits. The term "teacher characteristics" typically refers to qualities of teachers that can be measured with tests or derived from their academic or professional records. It does not generally refer to the direct observation of their impact on students' learning in terms of either students' test performance or teaching behaviors (both of which are addressed elsewhere in the present work). Rather, the approaches dealt with here are those that fall traditionally into the province of personnel psychology or personnel selection.[1] These

distinctions are particularly important because of the conclusions at which the present review arrives, namely, that psychometric selection is inappropriate in the initial selection of teachers and should defer to the evaluation of probationary teachers (teachers in the first few years of their employment).

RESEARCH ON TEACHER CHARACTERISTICS

MICRO-STUDIES AND MACRO-STUDIES

The research literature on teacher characteristics and student achievement encompasses two quite different kinds of study. One type – here referred to as *micro-studies* – uses individual teachers as the unit of analysis. Correlation coefficients are calculated from data descriptive of individual teachers and their students' achievement (usually expressed as a class average). Studies of this type yield findings most relevant to the question whether there are characteristics of teachers that predict their ability to improve the achievement of their students.

The second type of study is here called *macro-studies*. These studies measure characteristics of groups of teachers, such as "percentage of teachers in the school district who hold Masters degrees." Macro-studies attempt to exercise statistical controls by means of complex multiple regression analyses, often taking account of the multiple levels (states, districts, schools) of organization that tie individual teachers together. Macro-studies often inform policy at high levels but give limited direction to administrators who face individual selection decisions. Frequently, they do not express relationships in a form that permits the calculation of the actual benefits of selecting an elementary school teacher in terms of increased student achievement. Moreover, these

macro-studies – useful though they are for addressing state or national level policy questions – seldom achieve the levels of control needed to reach consensus among their readers. In spite of their contribution, macro-studies of the relationship between teacher characteristics as a school, district, or state "input" and student achievement as an "output" have several limitations: they must rely on imperfectly measured "background characteristics" of students to equate unequal conditions; they can not, without substantial and seldom realized extensions, resolve the ambiguity of the direction of the causal influence (Does a high percentage of Masters degrees raise student achievement, or do districts with able students who learn quickly and easily attract teachers with Masters degrees?); they typically fail to address the ambiguities present in ecological *correlation analysis*. (For instance, it is unclear whether the teachers holding the Masters degrees in the school district are the teachers actually responsible for the increased student achievement). Nevertheless, macro-studies of the relationship between teacher characteristics and student achievement are visible and influential at policy levels and will be reviewed here.

THE MICRO-STUDIES
Aptitude and Intelligence

Two major reviews of research[2] on the relationship of teachers' measured intelligence and their students' achievement arrived at the same conclusion: there is no important correlation between the two variables. Various explanations have been advanced for the failure to find a relationship that many expected would exist: the truncated variability of the intelligence scale for a population of teachers already highly selected for academic aptitude; the unreliability and lack of content validity of measures

of student achievement; as well as the essential irrelevance of high levels of measured intelligence for effective teaching, particularly at the elementary school level.

Academic Preparation

Research suggests that there is a modest relationship between teachers' college course work in the subject area in which they subsequently teach and their students' achievement.[3] Monk[4] analyzed data for almost 3,000 high school students from the Longitudinal Study of American Youth. Students took tests in mathematics and science, and supplied information on their backgrounds. Their math and science teachers were also questioned. Monk correlated teacher characteristics with student achievement, taking into account students' earlier achievement, background characteristics, and teacher inputs. The greater the number of college-level mathematics or science courses (or math or science teaching courses) teachers had taken, the better their students did on the mathematics and science tests. Goldhaber and Brewer[5] found similar relationships in a secondary analysis of more than 5,000 high school sophomores and their teachers. College-level math courses taken by the teachers was the only variable that accounted for any appreciable variation in students' achievement.

The National Teacher Examinations (NTE)

The National Teacher Examinations (NTE), developed and administered by the Educational Testing Service of Princeton, New Jersey, are widely used and an influential model for the state-level paper-and-pencil licensure exams that are currently proliferating throughout the United States. The validity of the NTE was the subject of an extensive review published in 1973 by Quirk, Witten, and Weinberg.[6] Subsequent reviews have not substantially added to nor altered their conclusions. Quirk *et al.* documented the nearly

30-year history of NTE research attempts to correlate NTE scores with such "concurrent validity" measures as high school GPA, undergraduate GPA, graduate GPA, ability tests (GRE-V, GRE-Q), as well as grades in specialized education courses. (Such correlations are referred to as "concurrent validity" coefficients because the two measures correlated are taken at roughly the same stage, in this case, during a prospective teacher's pre-service career.) Such criteria are only presumptively related to student learning; but even so, the concurrent validity evidence was not impressive. The highest correlations were with paper-and-pencil tests of academic ability and were in the region of 0.60. Paper-and-pencil tests correlate with other paper-and-pencil tests; that much might have been expected. Correlations of NTE scores with GPAs were in the region of 0.30. Most significantly, the two studies that produced correlations of NTE with grades in practice teaching yielded the following results: Shea[7] correlated NTE scores with grades in practice teaching for 110 pre-service teachers who had graduated from Worcester State Teachers College and obtained a r of –0.01; Walberg[8] correlated performance on the NTE with practice teaching grades for 280 pre-service teachers and found an r of –0.04. These are sobering findings indeed for those who hope for paper-and-pencil test information that will predict teaching effectiveness.

The usefulness of the NTE for predicting principals' ratings of various qualities of in-service teachers is similarly wanting. Research over 30 years in a wide variety of settings has shown correlations of NTE test scores and principals' ratings ranging from -0.15 to 0.50 with an average r of about 0.10.[9] In the face of these discouraging results, researchers have been prone to blame the professionals' evaluations of their peers and subordinates, suggesting that they are unreliable or biased or distorted by friendships or

prejudices or unsophisticated views of quality teaching. The fault, however, may lie more with the inadequacies of paper-and-pencil tests as measures if teachers' abilities to manage the complex demands of educating groups of children.

Quirk, Witte and Weinberg found only a single study in which NTE scores were correlated with students' average gain in performance from pretest to posttest, and this study by Lins,[10] published in 1946, produced data on only seven teachers. The correlation of NTE score with pupils' gain scores was 0.45; unfortunately, one can only assert with reasonable statistical confidence that a much larger sample would produce a validity coefficient somewhere between –0.50 and +0.90.[11]

The State of Massachusetts has instituted one of the most controversial paper-and-pencil teacher licensure tests. Haney and his colleagues found no empirical evidence that the Massachusetts teacher tests could predict student learning.[12]

Certification (Licensure)

A job candidate's certification status has become a visible consideration in recent decades as a result of a variety of reforms and economic pressures placed on the educational system. Class-size reduction efforts, most notably in California in the mid-1990s, not surprisingly created an acute need for teachers that could not be met by the existing supply of regularly certified personnel. The difficulty of recruiting certified teachers for schools in the deteriorating core of large cities prompted the hiring of college graduates without pre-service training or teaching experience – "Teach for America" being the most visible program of this type.[13] In addition, the market ideology that has influenced both the discussion and the implementation of education policy proposals since the 1980s questioned the need for state-operated systems of teacher certification.

Some believe that any educated person, with or without a college degree, can teach.[14] Educators are left with the question, what value is represented by the teacher license? Should certification status be considered in the hiring of new teachers?

Darling-Hammond wrote that "…reviews of research over the past thirty years, summarizing hundreds of studies, have concluded that even with the shortcomings of current teacher education and licensing, fully prepared and certified teaches are … more successful with students than teachers without this preparation."[15] Ashton[16] noted that teachers with regular state certification receive higher supervisor ratings and student achievement than teachers who do not meet standards, but this observation was based on data with virtually no statistical controls having been imposed. In spite of the quantity of research on the benefits of teacher certification for student learning that Darling-Hammond refers to, little of the past research exercised controls over student "inputs" that would give the critical reader confidence in the findings. One recent study addressed the effect of certification status with a series of controls that engendered this missing confidence.

Laczko and Berliner[17] studied the impact of certification status on student achievement in two large urban school districts. These school districts provided information about teachers hired for the 1998-1999 and 1999-2000 school years. Information included the school where they were currently teaching, the grade level taught, the teacher's certification status, highest degree earned, date and institution where it was achieved, age, and number of years teaching experience. Teachers were eliminated from the sample if they taught a grade level or subject that was not assessed (e.g., art and

music) by the Stanford Nine (SAT 9) achievement test battery, the measure of
achievement used in the study.

Emergency certified teachers were matched with regularly certified teachers in the
following manner: matches were first made by grade level; secondarily, matching was
based on highest degree attained; whenever possible, matches were made within the same
school, otherwise, matches were made within the same school district; cross-district
matching was not allowed. Matching the two samples produced 23 pairs of teachers for
the 1998-1999 school year and 29 pairs of teachers for the 1999-2000 school year.

Stanford Achievement Test-Version 9 scores aggregated at the class level for the
52 matched pairs of teachers were collected. Correlated t-tests were conducted to analyze
the difference in the student achievement scores between emergency certified and
standard certified teachers. The principal findings from the Laczko and Berliner study
appear in Table 8-1.

Table 8-1

**NCE Differences and Effect Sizes (ES)[18]
for 1998-1999 and 1999-2000
(After Laczko & Berliner, 2001)**

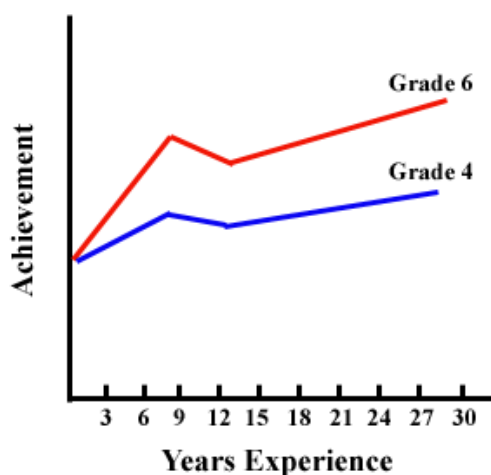| SAT 9 Sub-Test	1998-1999	1999-2000	Mean ES
Reading	13.9	9.2	0.50
Math	-0.2	11.1	0.24
Language	9.4	10.7	0.44

Note. The NCE differences are between certified and emergency teachers. Effect sizes
(ES) were calculated with a standard deviation of 23 NCE units.

Using the NCE (Normal Curve Equivalent) scale to express the results, Laczko and Berliner found, for example, that in the 1998-1999 school year, students taught by certified teachers outscored their counterparts taught by uncertified teachers by almost 14 NCE points in Reading. The similar margin in the 1999-2000 school year was greater than 9 points. Expressed as a proportion of the standard deviation of the NCE scale, these differences averaged across the two years yield an effect size of one-half (0.50) standard deviation (equivalent to five months grade-equivalent units). One would expect, based on these findings, then, that the students of certified teachers would make an additional five months academic growth in reading when compared to the students of uncertified teachers across an entire school year. The advantage for students of certified teachers in mathematics and language is one-quarter (0.25) standard deviation (about 2.5 months in grade-equivalents) and four-tenths (0.40) a standard deviation (about four months GE), respectively. These are, perhaps, the most convincing data yet produced by research on the effect of teacher certification on student achievement. (It should be noted that these differences in means expressed in standard deviation units correspond to correlations between certification status and student achievement of roughly 0.25, for effect sizes of 0.50, and 0.15, for effect sizes of 0.30 to 0.25.[19])

Successful Teachers of Poor Students

Poor students are disproportionately taught by less experienced teachers who are less likely to be licensed and who leave the profession sooner than teachers of the children of middle-class or wealthy families. Researchers have largely ignored the question of whether there are special characteristics of teachers who will be successful in teaching poor children.

One of the few quantitative studies of the relationship between teacher characteristics and student achievement for poor children is due to Murnane and Phillips.[20] Using data collected in a study of a federal welfare reform project in a large midwestern city, the researchers fit regression equations to account for the variability of vocabulary scores on the Iowa Test of Basic Skills in terms of teacher behaviors and other characteristics. The teachers were predominantly black, female and held Masters degrees. The researchers concluded: "Overall, the results … suggest that variables describing teacher behavior and variables describing teacher characteristics are both important in predicting teacher effectiveness."[21] Teacher characteristics of race, prestige of the undergraduate college, whether the teacher earned a Masters degree and verbal ability were not significantly related to students' achievement. However, "years of teaching experience" was related to student achievement. This relationship for Grades 4 and 6 is depicted in Figure 8-1. The relationship for Grade 1 was weaker, but still positive, and non-existent for Grade 5. No reasonable explanation for the interaction of the relationship with grade level exists, and a prudent conclusion would hold that teacher experience and student achievement are positively related in these circumstances.

Figure 8-1. Teacher Experience and Student Achievement for Inner-city Children (After Murnane, 1981).

Another one of the very few attempts to address this question was made by Martin Haberman in his book *Star Teachers of Children in Poverty.*[22] Drawing on years of interviewing hundreds of teachers in poor urban schools, Haberman advanced a view of what makes for success for a teacher of poor children. These successful teachers, which he named "star teachers," display the following characteristics: star teachers do not punish students, but instead use "logical consequences" to direct students to learn appropriate behaviors; star teachers believe that discipline problems are best handled by making learning interesting, meaningful, and engrossing; star teachers are persistent. Haberman saw these teachers dealing with the organization of the school in a uniquely productive way. They did not attempt to undermine the school's administration, nor did they ignore the directives of officials; however, they did not use bureaucratic directives as excuses to keep from achieving their objectives in the classroom. Star teachers engaged in what Haberman called "gentle teaching." Gentle teaching promotes kindness in classroom interactions; it pointedly avoids the discord that can characterize interactions in schools that emphasize compliance with rules instead of learning.

Haberman suggested that there may be ways to predict which teachers will be the star teachers. Candidates for teaching positions should be selected on the basis of criteria other than good grades and high test scores. New teachers, if they are to develop into Haberman's star teachers, should not be judgmental; they should be tolerant and avoid moralistic attitudes; they must be open, understanding, and not easily shocked; and they

must be capable of open and authentic communication with their superiors and colleagues.

Haberman has produced one of the few research-based works aimed at understanding the characteristics of teachers that make for success with poor children, and yet, his work has been criticized as methodologically weak.[23] No demographic description of the group of teachers interviewed is given; no explanation of the criteria by which the star teachers were recognized as successful is offered. Haberman may well be right, but the path traveled to reach his understandings is hidden from view.

THE MACRO-STUDIES

Large-scale studies that use school districts or states as the unit of analysis and attempt with multiple regression analysis to control for pre-existing differences among these units have addressed many of the same concerns analyzed in the micro-studies. The first large study of this type was Coleman's *Equality of Educational Opportunity*.[24] Coleman *et al.* measured seven characteristics of teachers: years of experience, highest degree attained, vocabulary test performance, ethnic group, parents' educational attainment, whether the teachers grew up where they were teaching, and the teacher's attitude toward teaching middle-class students. These teacher characteristics accounted for less than 1% of the variation in student achievement – meaning that a correlation of teacher characteristics with student achievement, holding other factors constant, would be less than +0.10. Coleman *et al.*, as well as Bowles and Levin,[25] felt that they detected slight relationships between teachers' verbal intelligence and student achievement. Summers and Wolfe[26] indicated that this relationship, though quite weak in statistical terms, was more important in some areas of the curriculum than in others. Hanushek[27]

joined these early researchers in finding no strong relationship between teacher characteristics and student achievement.

A pair of meta-analyses of macro-level studies arrived at differing conclusions on the question whether teachers' measured ability influences student achievement. Greenwald, Hedges, and Lane[28] reviewed a number of studies of the relationship between school inputs and student outcomes and concluded that teacher ability, teacher education, and teacher experience appeared to be related to student achievement. Hanushek's[29] synthesis of research studies arrived at a contrary conclusion regarding the relationship between teacher characteristics and student achievement. Less than a year later, Hanushek[30] published an "update" of his 1996 article in which he reported the following summary of studies that investigated the relationship (in terms of regression coefficients) between student achievement and their teacher's "years of experience."

Table 8-2

Direction and Statistical Significance of
Regression Coefficients for Student Achievement
Related to Teacher Experience
(After Hanushek, 1997)

		Stat. Signif.		Non. Signif.		
Ind. Var.	# of studies	+	–	+	–	?
Teacher Exper.	207	29%	5%	30%	24%	12%

Although a statistically significant regression coefficient for "teacher experience" was six times more likely to be positive than negative, Hanushek nonetheless read the results of Table 8-2 as negative for the effects of teacher experience on achievement. He wrote of the results: "A higher [than class size or teacher education] proportion of estimated effects of teacher experience are positive and statistically significant: 29%. Importantly, however, 71% still indicate worsening performance with experience or less confidence in any positive effect."[31] The logic of this conclusion is illusive. Of results that reach statistical significance, 85% (60/70) are positive, indicating that students of more experienced teachers achieve at higher levels. Of the statistically non-significant results *that can be determined*, 55% are positive, but fail to reach conventional levels of significance. Hanushek creates an impression of no effect of teacher experience by lumping together the category "indicative of worsening performance or less confidence of beneficial performance" all significant but negative coefficients (5%), all non-significant coefficients whether positive or negative (30% + 24%) and, remarkably, the 12% of the coefficients that were so incompletely reported that it could not be determined

whether they were positive or negative. The treatment of these data is hardly even-handed. By such logic, ten "positive studies," "no negative studies" and 100 studies so poorly reported that the results could not be discerned would lead to a conclusion of no confidence in a positive result. This author's reading of Table 2 is much different from Hanushek's. The data therein can be reasonably interpreted as evidence that regression studies have generally shown a positive relationship between teacher experience and student achievement.

Fetler[32] investigated the relationship between measures of mathematics teacher skill and student achievement in California high schools. Test scores are analyzed in relation to teacher experience and education and student demographics. The results are consistent with the hypothesis that there is a shortage of qualified mathematics teachers in California and that this shortage is associated with low student scores in mathematics. After controlling for poverty, teacher experience and preparation significantly predict test scores.

Darling-Hammond[33] utilized data from a survey of all 50 states' policies, the 1993-'94 Schools and Staffing Surveys of the U.S. Department of Education, and the National Assessment of Educational Progress to study the relationship between teacher qualifications and student achievement. The findings suggested that policy investments in the quality of teachers may be related to improvements in student performance. Measures of teacher preparation and certification were the strongest correlates of student achievement in reading and mathematics, both before and after controlling for student poverty and language status (limited English fluency v. full English fluency). "The most consistent highly significant predictor of student achievement in reading and mathematics

in each year tested is the proportion of well-qualified teachers in a state: those with full

certification and a major in the field they teach (*r* between 0.61 and 0.80, p<0.001). The

strongest, consistently negative predictors of student achievement, also significant in

almost all cases, are the proportions of new teachers who are uncertified (*r* between -0.40

and -0.63, p<0.05) and the proportions of teachers who hold less than a minor in the field

they teach (r between -0.33 and -0.56, p<0.05)." (It must be noted that these correlation

coefficients, in the area of 0.50 and above, are calculated on state-level aggregated data

and are much higher than would be obtained if similar variables were correlated at the

level of individual teachers.) Darling-Hammond's analyses suggest that state policies

regarding teacher education, licensing, hiring, and professional development may make

an important difference in the qualifications and capacities of teachers, and, as a

consequence, in the achievement of their students.

IMPLICATIONS FOR PERSONNEL SELECTION
Correlations and Base Rates

It is common in research on the relationship of teacher characteristics and student

achievement to express the relationship in terms of correlation coefficients. Such

coefficients have distinct disadvantages in communicating the benefits of selecting

teachers on the basis of their entry characteristics (such as college GPA, NTE scores,

scores on teacher certification exams, Teacher Perceiver profiles and other similar

measures of potential). Correlations of beginning teacher characteristics and their

students' eventual achievement are typically in the range of 0.15 to 0.35, as was seen in

the research reviewed above. The lay reader is frequently misled into thinking that such

relationships possess a practical benefit when the finding is referred to as "statistically

significant." This may not and – in the present application of psychometrics – probably is not the case. "Statistical significance" is a quality of statistical findings that refers only to their reliability or "inferential stability," that is, the likelihood that a particular finding has not arisen by chance sampling from a population in which the two variables correlated are completely unrelated. Statistical significance results from taking large samples, and generally means nothing more than that the statistical finding was based on a large sample. The finding itself could be of no practical value and still be "statistically significant." Persons' heights and their IQs might correlate 0.02 in a sample of 100,000 persons and be deemed "statistically significant"; but that finding will be of no value whatsoever.[34]

The benefits, if there are any, of selecting teachers on the basis of such weak correlational evidence – validity coefficients in the range of 0.35 and below – are not clearly seen in correlation coefficients. The meaning of these relationships is more clearly seen in statistics such as "hit rates" or measures of "false positives" and "false negatives" – for example, the differences in percentages of teachers who will not survive their probationary evaluation between those who score high on some characteristic, such as college GPA, and those who score low on that characteristic.

Consider what will prove to be a typical situation: the district's assistant superintendent for personnel has available the college GPA of all applicants for openings in elementary education. There are twice as many applicants as there are openings, so she selects the top half of the applicants on the basis of their GPA. Suppose further that the correlation between teaching candidates' GPA and their students' learning is 0.35 – a not unreasonable assumption, surely not an underestimate. Furthermore, suppose that 5% of

the probationary teachers in this district are not rehired after two years and that the rehire decision is based solely on their ability to engender student learning.[35]

Table 8-3

**Hypothetical Relationship Between
Selection Criterion and Success Criterion**

	Re-Hired	Not re-hired	Totals
Selected	490	10	500
Rejected	460	40	500
Totals	950	50	1,000

Table 8-3 shows counts of teaching candidates selected or rejected on the basis of their college GPAs and the result of the decision to continue employment after their probationary period. The data in Table 8-2 correspond to a correlation of GPA and "teaching success" of approximately 0.35 with a selection rate of 50% and a success rate of 5%. Meehl and Rosen[36] pointed out nearly 50 years ago that the utility of a correlation in predicting an event (like success in teaching as evidenced by continuing employment) depends on: a) the size of the correlation, b) the costs of errors in prediction (of rejecting a person who would succeed or accepting a person who will eventually fail), and c) the "base rate" of the event being predicted. (Also see Wainer's application of these concepts to the Massachusetts Teacher Tests).[37] The major implication of Meehl and Rosen's argument is this: if the event being predicted has a very low incidence of occurring (a "low base rate"), then very large correlations of predictors with the criterion are needed or else one makes fewer errors by using no predictor whatsoever.

One can see this phenomenon at work in the above table. If teaching candidates are selected because they have high (top half) GPAs, 10 out of 500 candidates will not be re-hired, and 460 out of 500 who would have succeeded if they had been hired will never get a chance to show that they could have succeeded. Applicants with a high GPA (and who are selected) have a 2% probability of "failing" (i.e., not being rehired). But applicants with a low GPA (who would not have been selected) have only an 8% probability of failing (i.e., not surviving the probationary period). The use of the GPA in selecting new teachers represents a gain in detecting "success" of from only 2% to 8%, but this gain comes at the cost of rejecting 92% of new hires would eventually would prove to be successful. In most people's system of values, rejecting 92% of potentially successful applicants in order to achieve a 98% success ratio in prediction is unfair to a large number of applicants. Psychometricians say that in these circumstances the cost of "false negatives" is too high.

Furthermore, when an administrator can control the overall rate of "success" (say, for example, when 95% of teachers receive "merit pay" bonuses and the discretion exists to raise that rate to 100%), it is frequently the case that even a good predictor of that 95% will create more erroneous decisions than declaring all 100% of teachers successful, hence using no selection criterion at all. Validity coefficients are not sufficient for evaluating the practical utility of a test or other selection technique: "... when the base rates of the criterion classification deviate greatly from a 50 percent split, use of a test sign having slight or moderate validity will result in an *increase* of erroneous clinical decisions."[38]

Between and Within District Variation

A second problem exists in translating the research on teacher characteristics into the real world of personnel decisions. In research studies, an effort is made to sample a full range of subjects (persons) along the continuum of the characteristics being correlated with student learning. But in the real world of schools, teacher applicants and students are clustered into schools and districts that represent selected portions of these continua. It may often be the case that a teacher characteristic that has shown modest correlations with student achievement in research studies will have no relationship with achievement *within* the particular school district attempting to select the best teachers for its students. This possibility – which is a highly likely circumstance – is illustrated in Figure 8-2.
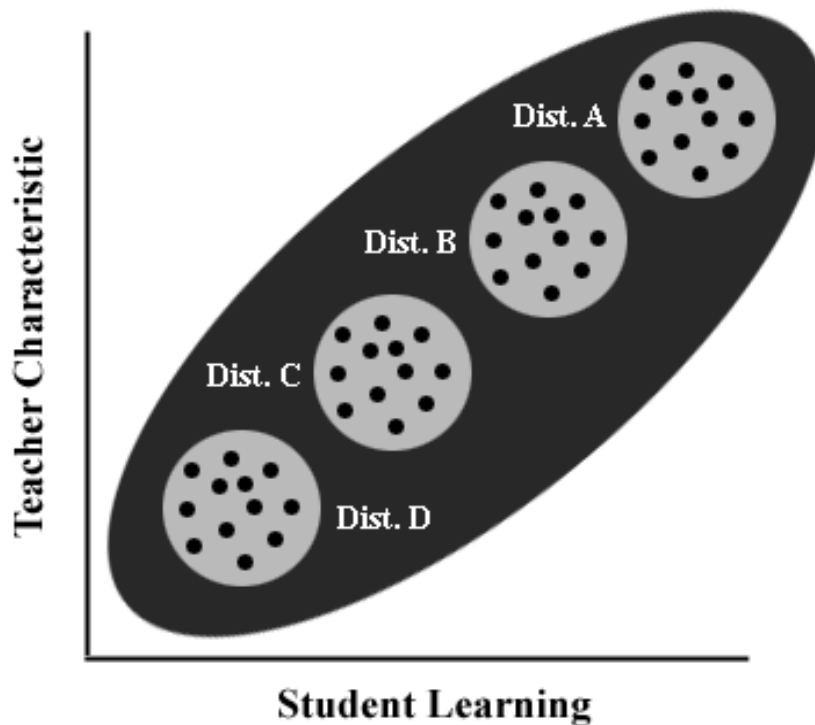

**Figure 8-2. Illustration of Between and Within School District
Relationships of a Teacher Characteristic and Student Learning**

Figure 8-2 illustrates a hypothetical situation in which 12 teachers are measured in each of four school districts on a characteristic (such as college GPA, for example) and on their contribution to their students' learning. It should be noted that the degree of relationship between a teacher characteristic and student learning depicted in Figure 8-2 is far greater than anything ever demonstrated in an actual research study, but this exaggeration will strengthen rather than vitiate the point being illustrated. Within each school district there is zero correlation between the measured teacher characteristic and the students' learning; however among the four districts, the teacher characteristic and student learning are highly correlated, perhaps as high as a coefficient of 0.80. The import of this situation is significant, however. What this arrangement of variation between and within districts implies is that the teacher characteristic is of no use whatsoever for selecting teachers within any one school district. And since it is within particular school districts that administrators live and work, knowledge of the teacher characteristic is of no value to them in selecting teachers who will enhance their students' learning.

This point may appear to be simply argumentative and counter-intuitive. The implication of this observation is real, however, and not simply some statistical sleight of hand. It dampens enthusiasm for the meager correlations that have been found; and coupled with the earlier observation on the relationship between correlation coefficients and hit ratios, it underlies the ultimate recommendation made here on the matter of initial teacher selection.

Finally, one more point must be raised that will further temper one's expectations of finding here clear statistical evidence for selecting teachers who can promote student

learning. A proper predictive validity study would involve randomly assigning students to groups (or some careful matching of students across groups to ensure their initial equivalence), then randomly assigning groups to teachers, measuring teacher characteristics, allowing instruction to proceed for some substantial period, measuring student learning, and then correlating the groups' learning gains with the teacher characteristics for many teachers. It would be crucial to measure student learning by means of their *gains* in performance from before to after instruction. Simply to correlate teacher characteristics with students' achievement, as has been done repeatedly in the research literature, would not accomplish the purpose of relating teacher characteristics to student learning. Because of the many factors that influence which teachers are employed in which schools in the world outside the research laboratory – teachers with higher GPAs, and measured aptitude, perhaps, are employed in schools whose students enjoy many advantages over schools that face the challenges of poverty and discrimination – the correlation of teacher characteristics with (uncorrected) student achievement test scores measures little more than the often remarked upon sorting of more able teachers into privileged schools. Nothing like this research has ever been published, in part because of the obvious expense, the impracticality of arbitrarily constituting actual school classes of students and randomly assigning them to teachers, and, perhaps, because of researchers sense that the payoff in terms of useful predictive information would be meager. (The "micro-teaching" studies of the 1960s and early 1970s at Stanford University approximate this ideal design in terms of controls, but the focus there was on teacher behaviors that promote student learning.) A thorough literature review in the preparation of the current work revealed a single study that even approached the

conditions stated above for a proper study, and that study[39] was published more than 50 years ago.

SUMMARY AND RECOMMENDATIONS

The early promise of psychometric techniques for the initial selection of teachers seems to have all but disappeared from the agenda of researchers; it may never have held a prominent place in the actual practice of educators.[40] Though rare exceptions can be found (e.g., the Montgomery County, Va., schools in the 1980s, as described by Wise *et al.*[41]), actual selection of teachers in America's schools is today based on interviews and personal interactions that reveal evidence of the candidate's appearance, enthusiasm, personal style and similar attributes. Measurement of ability, past achievements, or the candidate's ability to produce learning gains for students plays virtually no role in the selection of new teachers. This is not to say that the current practice is to be disapproved of. Current practice in teacher selection probably reflects an understanding that the cohesiveness of a school's staff is more critical to the success of the school and its students than is the level of teachers' performance on paper-and-pencil tests of dubious validity.

The customary procedure for selecting new teachers is based more often on first-hand experience with the candidate's teaching than it is on psychometric evidence in the form of test scores, GPAs or other evidence of personal characteristics believed to be predictive of successful teaching.[42] Schools often choose their new teachers from among interns and student teachers for whom the teaching staff has direct knowledge of their teaching abilities. Alternatively, substitute teachers are observed and evaluated as potential candidates. The arguments marshaled here against psychometric selection of

new teachers, because of low correlations of teacher characteristics with student learning and very low base rates of releasing probationary teachers, have already worked their way into the existing system of evaluating candidates for new hires. The need is not for better instruments to measure initial teachers' aptitudes and dispositions, but for better methods of evaluating more directly the ability of probationary teachers to foster learning in their students.

The measurement of the direct contribution that a teacher makes to the learning of his or her students is an enormously difficult technical problem that, in the opinion of the author, has no adequate solution that can be applied with confidence under real world conditions. The attempt to base teachers' rewards (salary increases, for example) on measured student progress is even more problematic,[43] as is noted elsewhere in this report.

The claim that psychometric measures of teacher characteristics are not useful for initial teacher selection implies that candidates be selected by other means – staff interviews, recommendations by peers or past supervisors, and the like. Some might think that this approach is an abrogation of responsibility; but instead, it is a realization of the limits of psychometric approaches to personnel selection. The true abrogation of responsibility is when professional educators – whether they are tenured teachers, administrators or professors engaged in pre-service education of teachers – fail to conduct adequate evaluations of pre-service and in-service teachers who are practicing their profession under the supervision of their superiors.

These findings, then, yield the following recommendations:

- Paper-and-pencil tests are not useful predictors of teaching candidates' potential to teach successfully and should not be used as such.

- Teaching candidates' academic record (e.g., GPA) is not a useful predictor of their eventual success as teachers. A candidate's record of success in pre-service (undergraduate) technical courses (mathematics and science, for example) may contain useful information about that candidate's success in teaching secondary school mathematics and science.

- Other things equal, 1) students of regularly licensed teachers achieve at higher levels than students of emergency certified teachers; and 2) more experienced teachers produce higher student achievement than less experienced teachers. Teacher selection policies should reflect these facts.

- The selection of teachers who will best contribute to their students' academic achievement should focus on peer and supervisor evaluation of interns, student teachers, substitute teachers and teachers during their probationary period.

REFERENCES

1 L. J. Cronbach and G. C. Gleser, *Psychological Tests and Personnel Decisions*, 2nd ed. (Urbana: University of Illinois Press, 1965).

2 D. Schalock, "Research on Teacher Selection," in *Review of Research in Education, Vol. 7,* ed. D. C. Berliner (Washington, D.C.: American Educational Research Association., 1979).

   R. S. Soar, D. M. Medley, and H. Coker, "Teacher Evaluation: A Critique of Currently Used Methods," *Phi Delta Kappan* 65, no. 4 (1983): 239-246.

3 C. A. Druva and R. D. Anderson, "Science Teacher Characteristics by Teacher Behavior and by Student Outcome: A Meta-Analysis of Research," *Journal of Research in Science Teaching* 20, no. 5 (1983): 467-479.

   See also: V. A. Perkes, "Junior High School Science Teacher Preparation, Teaching Behavior, and Student Achievement," *Journal of Research in Science Teaching* 6, no. 4 (1967-1968): 121-126.

4 D. H. Monk, "Subject Matter Preparation of Secondary Mathematics and Science Teachers and Student Achievement," *Economics of Education Review* 13, no. 2 (1994): 125-145.

5 D. D. Goldhaber and D. J. Brewer, "Why Don't Schools and Teachers Seem to Matter? Assessing the Impact of Unobservables on Educational Productivity," *Journal of Human Resources* 32, no. 3 (1996): 505-520.

6 T. J. Quirk, B. J. Witten, and S. F. Weinberg, "Review of Studies of Concurrent and Predictive Validity of the National Teacher Examinations," *Review of Educational Research* 43 (1973): 89-114.

7 J. A. Shea, *The Predictive Value of Various Combinations of Standardized Tests and Subtests for Prognosis of Teaching Efficiency* (Washington, D.C.: Catholic University of America Press, 1955).

8 H. J. Walberg, "Scholastic Aptitude, the National Teacher Examinations, and Teaching Success," *Journal of Educational Research* 61 (1967): 129-131.

9 Quirk *et al*, Table 2.

10 L. Lins, "The Prediction of Teaching Efficiency," *Journal of Experimental Education* 15 (1946): 2-60.

11 G. V Glass and K. D. Hopkins, *Statistical Methods in Education and Psychology*, 3rd ed. (Boston: Allyn & Bacon, 1996), 357.

12 W. Haney et al., "Less Truth Than Error? An Independent Study Of The Massachusetts Teacher Tests," *Education Policy Analysis Archives* 7, no. 4 (1999), <http://epaa.asu.edu/epaa/v7n4/.>.

13 W. Kopp, "Teach for America: Moving Beyond the Debate," *The Educational Forum* 58, no. 4 (1994): 187-192.

14 G. W. McDiarmid and S. Wilson, "An Exploration of the Subject Matter Knowledge of Alternative Route Teachers: Can We Assume They Know Their Subject?" *Journal of Teacher Education* 42, no. 2 (1991): 93-103.

15 L. Darling-Hammond, *The Right to Learn: A Blueprint for Creating Schools That Work* (San Francisco, CA. Jossey-Bass, 1997), 308

16 P. Ashton, "Improving the Preparation of Teachers," *Educational Researcher* 25, no. 9 (1996): 21-22.

17 I. I. Laczko and D. C. Berliner, "The Effects of Teacher Certification on Student Achievement: An Analysis of the Stanford Nine," paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA, 2001.

18 The practical significance of a study is often expressed in a form known as an *effect size*. An effect size that measures the amount of difference between two groups is defined as a mean difference (between conditions A and B) in units of the within-condition standard deviation:

ES = (Mean-A − Mean-B) / σ

The value of *ES* reveals the amount of superiority of condition A over condition B (or, B over A in the event that *ES* has a negative value). Under the assumption of normally distributed scores, an *ES* of +1.0 indicates that the average student in condition A scores above 84% of the students in condition B. When the effect size is calculated on standardized achievement test data, a fortuitous coincidence gives the measure added meaning. It is an empirical fact that the standard deviation of most achievement tests is 1.0 years in *grade equivalent units*. Consequently, an effect size of 1.0 implies that the average superiority of condition A over condition B is 1.0 in grade equivalent units. Likewise, an effect size of .50 implies that students in A achieve, on average, 5 months in grade equivalent units above students in condition B.

19 Glass and Hopkins, 1996

20 R. J. Murnane and B. R. Phillips, "What Do Effective Teachers Of Inner-City Children Have In Common?" *Social Science Research* 10, no. 1 (1981): 83-100.

21 Ibid., 91.

22 M. Haberman, *Star Teachers of Children in Poverty* (Bloomington, IN: Kappa Delta Pi, 1995).

23 E. L. Brown, review of *Star Teachers of Children in Poverty*, by Martin Haberman, *Education Review* (July 22, 1999), <http://coe.asu.edu/edrev/reviews/rev64.htm>.

24 J. S. Coleman et al., *Equality of Educational Opportunity* (Washington, DC: U.S. Government Printing Office, 1966).

25 S. Bowles and H. M. Levin, "The Determinants of Scholastic Achievement – An Appraisal of Some Recent Evidence," *Journal of Human Resources* 3 (1968): 3-24.

26 A. A. Summers and B. L. and Wolfe, *Which School Resources Help Learning? Efficiency and Equality in Philadelphia Public Schools* (Philadelphia, PA: ED 102, February 1975), 716.

A. A. Summers, and B. L. Wolfe, "Do Schools Make a Difference?" *American Economic Review* 67 (September 1977): 639-652.

27 E. A. Hanushek, "Teacher Characteristics and Gains in Student Achievement: Estimation Using Micro Data," *The American Economic Review* 61, no. 2 (1971): 280-288.

28 R. Greenwald, L. V. Hedges, and R. D. Laine, "The Effect of School Resources on Student Achievement," *Review of Educational Research* 66 (1996): 361-396.

29 E. Hanushek, "A More Complete Picture of School Resource Policies," *Review of Educational Research* 66, no. 3 (1996): 397-409.

30 E. Hanushek, "Assessing the Effects of School Resources on Student Performance: An Update," *Educational Evaluation and Policy Analysis* 19, no. 2 (1997): 141-164.

31 Hanushek, 144.

32 M. Fetler, "High School Staff Characteristics and Mathematics Test Results," *Education Policy Analysis Archives,* 7, no. 9 (1999), <http://epaa.asu.edu/epaa/v7n9.html>.

But also see:

E. J. Fuller, *Does Teacher Certification Matter? A Comparison Of TAAS Performance in 1997 Between Schools with Low and High Percentages of Certified Teachers* (Austin, TX: Charles A. Dana Center, University of Texas at Austin, 1999).

L. Darling-Hammond, "Teaching and Knowledge: Policy Issues Posed by Alternate Certification for Teachers," *Peabody Journal of Education* 67, no. 1 (1990): 123-154.

33 L. Darling-Hammond, "Teacher Quality and Student Achievement: A Review of State Policy Evidence," *Education Policy Analysis Archives* 8, no. 1 (2000), <http://epaa.asu.edu/epaa/v8n1/>.

34 Glass and Hopkins, 269.

35 This estimate is actually higher than the prevailing figures in public schools. An informal survey of teacher educators and administrators conducted in June 2001 on the AERA Division K listserv fixes the true figure at 5% or less.

36 P. E. Meehl and A. Rosen, "Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores," *Psychological Bulletin* 52 (1955): 194-216.

37 H. Wainer, "Some Comments on the Ad Hoc Committee's Critique of the Massachusetts Teacher Tests," *Education Policy Analysis Archives* 7, no. 5 (1999), <http://epaa.asu.edu/epaa/v7n5.html>.

38 Meehl and Rosen, 215.

39 Lins.

40 W. Haney, G. Madaus, and A. Kreitzer, "Charms Talismanic: Testing Teachers for the Improvement of American Education," in *Review of Research in Education*, vol. 14., ed. E. Z. Rothkopf, (Washington, D.C.: American Educational Research Association, 1987), 169-238.

41 A. E. Wise et al., *Effective Teacher Selection: From Recruitment to Retention – Case Studies* (Santa Monica, CA: The RAND Corporation, 1987).

42 Ibid.

L. Darling-Hammond, A. E. Wise, and S. R. Pease, "Teacher Evaluation in the Organizational Context: A Review of the Literature," *Review of EducationalRresearch* 53 (1983): 285-337.

43 G. V Glass, "Using Student Test Scores To Evaluate Teachers," in *New Handbook of Teacher Evaluation*, eds. J. Millman, and L. Darling-Hammond (Beverly Hills, CA: SAGE, 1989).