DO GRADES AND TESTS PREDICT ADULT ACCOMPLISHMENT?

Leonard L. Baird

The relationship of measures of academic ability and grades with high level accomplishment was examined by reviewing a wide ranging literature. This literature included studies of the highly creative, scientists and technicians, physicians, high- and middle-level managers, and high school and college students. The Terman studies of the gifted were also reviewed. Finally, studies of occupational attainment and income were examined. A very wide variety of criteria were used in these studies. In general, the studies demonstrated low positive relationships between academic aptitude and/or grades and accomplishment. The closer the content of the measure of academic aptitude to the demands of the field, the stronger the relationship.

Skepticism about the value of academic ability and academic success has grown in recent years. Increasing numbers of researchers, professional psychologists, and laymen have questioned the importance of high grades and high scores on academic ability tests. Doubts have been greatest about the role of academic talent in high-level or creative accomplishment. The purpose of this review is to examine the research evidence about the relationship between measures of academic ability and high-level accomplishment. It is important to understand the relationship because admission to many colleges is based primarily on academic ability. Academic ability is also a prime consideration in the award of scholarships and financial aid. Finally, academic ability and success are often considerations in hiring people for jobs in industry, education, and government. To justify these practices, it needs to be shown that selection on academic ability leads to the choice of people of above average potential who will later contribute to their fields or positions.

Why should we expect a relationship between academic ability and/or grades and real-life accomplishment? Although this is a straightforward

Leonard L. Baird, Department of Higher Education, University of Kentucky.

question, there are very few direct answers to it. Even writers such as Herrnstein in *I.Q. in the Meritocracy* do not discuss the logic behind the relationship, although they discuss a good many studies. More often than not the answers or logic are presumed and not spelled out. Perhaps the most basic assumption is that academic ability plays a large or at least contributing role in success in most human activities. The ordinary person might express this as "you've got to be smart to do a really good job or to get ahead"—with "smart" usually meaning that a person would do well in school or would have skills that would lead him or her to score high on a test of academic ability. In fact, Duncan, Featherman, and Duncan (1972) obtained a correlation of .91 between the prestige ratings of 47 occupations and ratings of the amount of intelligence each occupation was believed to demand. This belief may be reasonable, since most attainments are to some degree dependent on the ability to read, understand, and analyze written materials, and on knowledge and understanding of mathematical concepts such as those represented in academic ability tests and the classroom. An executive who rises rapidly in a company, an engineer who files a patent, or a writer who publishes a story are usually thought to be "smart" in the sense just described. Put more formally, it is assumed that high-level accomplishment includes intellectual demands for its attainment that require a fairly high level of academic ability. The person with high academic talent should thus be able to attain more than the person with little academic talent.

To a large degree this idea is implicit in grading in schools and colleges. The student who does well in the classroom is expected to be able to do well in real-life situations. The classes and curriculums are designed to prepare students to function as citizens and workers in the general society and in specific occupations and professions. Thus, the students who do well in class should also generally do well in the social roles and occupational duties for which these classes have prepared them.

From this it follows that the students who will be most likely to succeed in society and in particular occupations and professions are those who have the most academic talent and who have had the greatest success in academic work. Consequently, admission to colleges and professional schools is based primarily on measures of academic ability and records of previous grades. Again, the basic assumption of admissions policies is that the students with the greatest academic ability and accomplishment are the most likely to do well in coursework, and consequently in society and in their occupations. Likewise, selection for many jobs is also based to a large degree on measures or records of academic talents. Thus the assumption that academic ability is an important, perhaps even required, element in human accomplishment is reflected in the popular culture, educational system admission practices, and employment selection.

Related to this pervasive view of the role of academic talent is the meritocratic belief that academic talent should be a primary reason for admission to schools and careers and for advancement in those areas. That is, it is much more preferable that admission and promotion decisions be based on academic ability rather than on the accidents of family wealth, ethnic background, religion, or neighborhood. This policy seems fairer to individuals, encourages healthy competition, and results in the most able holding positions of responsibility in our society. Again this belief is based on the idea that academic ability is important in successful functioning in high-level roles in our society.

Why should we expect to find little relationship between academic ability and real-life accomplishment? Although the idea that academic ability is important, in many if not most high-level accomplishments, is pervasive and persuasive, there are arguments against finding such a relationship. The first stems from the fact of the sheer diversity and specificity of human activities. It seems unreasonable to expect academic ability to be highly related to success in such divergent areas as management, leadership, community service, religion, music, technical work, scientific research, artistic work, literary work, dramatic activity, journalism, etc. Experts within each of these areas show even greater differentiation. For example, literary work in different areas is said to involve very different skills (e.g., writing short stories is different from writing topical magazine articles, which is different from writing novels, which is different from writing books on factual matters, which is different from writing scholarly books, etc.). Editing these different types of writing also involves different kinds of skills, as do producing them and publishing and promoting them.

As this last example suggests, human activities are also situationally specific. An executive may do quite well in one company and poorly in another or may do well or poorly even in the same company depending on the details of his or her position, the outcome of a few key projects, the quality of subordinates' work, or the character of superiors. There are many stories of scientific discoveries that were dependent as much on accident as on the ability of the scientist. In many cases, accomplishment may be due to the right person being in the right place at the right time. Thus, given equally able and equally trained people, accomplishment may be dependent on the specific situations people find themselves in. In sum, people do so many things in so many contexts, that it may be unreasonable to expect academic ability to be highly related to attainment in every situation.

Another reason for expecting little relationship between academic ability and high-level accomplishment lies in the definitions and criteria of accomplishment. Accomplishment in many areas is a complex, multifaceted thing. For example, Crooks and Campbell (1974) used over 30 measures as criteria

of managerial success, and found that they were only moderately intercorrelated. Taylor and his associates used 77 measures to define physician performance (Price et al., 1973). Again the measures and factors were not highly interrelated. Likewise, many criteria of success have ambiguous meanings. For example, annual income has been used as a criterion of the overall success of college graduates. This criterion might appear to be clear, objective, and applicable to everyone. However, many professions such as the clergy and elementary school teaching are low paying; conversely, having a high income may be due to some lucky investments, inheritances, working in high-profit businesses such as oil, or positions in family businesses. Furthermore, it would be difficult to argue that a lawyer serving most of the citizens of a small town who earned $20,000 a year was less "successful" than a classmate who worked in bond transactions on Wall Street and earned $100,000. Other accomplishments are so rare as to call into question their applicability across a profession. For example, several surveys of Ph.D. psychologists have found that most psychologists have never published an article. Even within colleges and universities, Ladd and Lipset (1975) found that 29% of faculty members had never published an article (nearly half had published fewer than three) and 59% had never published a book (86% had published fewer than three). In contrast, only 9% had published 31 or more articles and only 6% had published five or more books.

Other researchers have argued that the idea of a single dominant talent, such as that represented in academic ability, is wrong. They contend that there are many other human capabilities that play a role in human performance. Thurstone (1938), for example, used factor analysis to identify seven basic aptitudes. The United States Employment Service factor-analyzed a vast number of tests and located and measured nine factors which were then related to the requirements for occupations in the Dictionary of Occupational Titles. Finally, Guilford (1968) has proposed a model of the "structure of the intellect" that includes 120 different factors, the majority of which he claims to have demonstrated to exist. Whatever is the most accurate way to describe human abilities, it is clear that academic capacity is only part of the possible range of abilities. Whether it plays the dominant role in human accomplishment is a matter of debate. In any case, in those situations where other abilities play a large role, it may be difficult to show the independent effect of academic ability.

Another limitation of the studies reviewed here is the length of time between the assessment of academic talent and the assessment of accomplishment. For example, an investigator may attempt to relate grades in college to income or performance in a profession 10 or 15 years later. Obviously many things can happen in 10 or 15 years to affect an individual's career. The choice of employers, region, and spouse can have an influence as

can accidents, sickness, and personal problems. There are many nonacademic influences even within a specialized profession. For example, a physician's career can be influenced by his or her choice of specialization, area of practice, hospital or laboratory, partners, etc. Thus, the longer the time following the assessment of academic talent, the greater role life circumstances can play, and the lower the relationship with accomplishment. Of course, if it is true that once people are shown to be "smart" in school or college, then they will continue to make "smart" decisions throughout their lives and generally perform better than other people, it follows that life circumstances will not have a very pronounced effect on their careers.

Another, more technical reason for expecting little demonstrable relationship between academic ability and accomplishment is the statistical inadequacy and unreliability of the criteria. As suggested earlier, the distributions of many criteria of accomplishment may be highly skewed; others are categorical, such as winning a professional prize or not; and still others represent a summation of very different behaviors, so their meaning and stability is questionable. Many criteria are quite unreliable. For example, "superior ratings" are sometimes based on the ratings of only a single person or are summations of ratings of people who have very different degrees of experience with the individual being rated. Other criteria are based on inadequate records and other sources of data that adversely affect their reliability. The important point is that the lower the reliability of the criterion, the lower the degree to which it can be predicted. In many of the studies reviewed here, the criteria have moderate reliability at best, limiting the degree to which academic ability can be shown to be related.

Another technical limitation is the range of academic ability present in a study sample. The narrower the range of academic talent, the lower the relationship that can be demonstrated between academic ability and the criterion. Consider, for example, a study of the career successes of Phi Beta Kappa recipients. Since the students all had extremely high grades it would be difficult to distinguish between very successful and less successful students on the basis of grades. In contrast, a study of the career successes of an entire college class would allow a broad range of academic talent to be studied, and students could be distinguished on the basis of grades. Thus, to be able to show a very strong relationship between academic ability or grades and accomplishment, a reasonably wide range of academic talent is required, something that is seldom the case in the studies reviewed here.

The effects of restriction of range on the correlation can be quite dramatic. For example, if the actual correlation across a total group were .30 and only the top half of the distribution were selected, then the correlation found would be .1, and if only the top 10% were selected, then the correlation found would be .12. Most professionals have been selected for college,

graduated from college, admitted to professional school, and graduated from professional school, all largely on the basis of their academic performance. Thus, those who enter a profession have already been selected for academic talent several times, each time at a higher level. The result is that most professions include a rather narrow range of academic talent. Similarly, many companies select their staff at least partly on the basis of their academic record; some also use their own ability test. The result here is also a narrower range of academic talent.

Another technical factor limiting the demonstrated relationship between academic ability and accomplishment is the small size of many study samples and the limitations of statistical tests for demonstrating relationships. This is due to a basic element of statistical procedures, that their effectiveness is dependent on the size of the sample. A statistical procedure has more "power" to correctly detect a relationship as the sample increases. With small samples this power is quite limited, so that true relationships may not be detected, and thus the incorrect conclusion is reached that there is no relationship. Many of the studies of the relationship between academic talent and success are based on small samples, so that the studies may conclude that there is no relationship when in fact there is one. (The concept of power, as used here, is somewhat technical; the reader is referred to Trattner and O'Leary (1980). For example, if the actual correlations is .30, a researcher would correctly detect a significant correlation ($p < .05$) only 54% of the time with an N of 50, and 85% of the time with an N of 100.)

Another technical problem is that admissions decisions tend to be compensatory (Dawes, 1971). That is, when an applicant is low on one admissions measure, he or she may be admitted on the basis of high standing on another. For example, applicants to graduate school who have very low undergraduate grades will be admitted only if they have very high aptitude test scores, and vice versa. If these individuals are then successful in graduate work, any study relating their undergraduate grades or test scores to success may show a small or even a negative relationship. This may be especially critical when the criterion is some sort of creative or high-level accomplishment. For example, applicants with both low undergraduate grades and low test scores may be admitted to a chemistry program if they have already shown signs of scientific promise (e.g., by publishing articles on their own original research). Since the best predictor of later high-level accomplishment is earlier accomplishment in the same area, a study of the relationship between grades, aptitude test scores, and the scientific achievements of the graduates of the chemistry program may show little relationship.

Finally an argument can be made that to the extent education is successful in bringing students up to a common standard of competence, the more

difficult it is to demonstrate a relationship between academic ability and accomplishment. That is, when an educational program—for example, in a profession like medicine—effectively prepares all its students for the demands of a particular occupation or profession, it "equalizes" the differences in the academic ability of the students. If all the students are prepared to meet successfully the intellectual and personal demands of occupational or professional work, then the differences in the degree of their success are likely to be due not to their academic ability to master an academic program but to other factors, such as specific skills and personal characteristics. Of course, not all programs can be completely effective, so the possible influence of academic talent is still present; however, to the degree they are effective, the more difficult it is to demonstrate its potential role.

In sum, there are conceptual and technical reasons to expect to find little relationship between academic ability and accomplishment. Whether the influences that would limit the degree of the relationship are strong enough to mask the kind of relationship expected, because of the reasons discussed earlier, is an important concern of this review. Dissentangling the limiting influences to reach an estimate of the true relationship is the chief work of the review.

EARLIER REVIEWS OF THE RELATIONSHIP BETWEEN ACADEMIC PERFORMANCE OR TESTS OF ACADEMIC ABILITY AND ADULT ACCOMPLISHMENT

The most comprehensive review of studies of the relationship between college grades and adult accomplishments conducted prior to 1965 was written by Hoyt (1966). Hoyt noted that many of the 46 studies he examined were flawed by sampling bias, difficulties with criteria, and measurement problems. Nevertheless, Hoyt concluded that the correlation between grades and adult accomplishment was very small, and in many cases near zero. Nelson (1975), in a government report for the Civil Service Commission, discussed more recent studies, but reached a similar conclusion. O'Leary (1980), in another government report for the Office of Personnel Management, summarized the relationship between grades and occupational success found in 28 studies. Weighting the reported correlations by the size of the sample, O'Leary found an average correlation of .17.

Two recent studies used Glass' (1978) method of meta-analysis to examine the relationship between college grades and adult achievement as indicated by income, ratings, etc. Samson et al. (1984) examined 35 studies, and the influence of such variables as the time of measurement of the outcome variable, gender, major field in college, etc. Although noting the influence of such factors as the homogeneity of the samples and the invalidity of meas-

ures of occupational success, they conclude that "Even so, the overall variance accounted for makes academic grades or test scores nearly useless in predicting occupational effectiveness and satisfaction." The study-weighted mean correlation was .155, accounting for 2.4% of the variance.

Cohen (1984) used the same methodology, but used 108 studies and examined the influence of the characteristics of the studies such as the selectivity of the institutions in the studies, the year the study was published, etc. The average correlation between grades and overall success was .18, and the correlations for more specific criteria of adult success ranged between .09 and .20. These criteria included ratings, income, promotions, eminence, etc. He concludes: "The results presented in this paper may be somewhat discouraging to those placing great importance on grades and their predictive power. It seems that how well a student does in college relates only marginally with success in a career" (p. 292).

Although not actually a review, the work of Thorndike and Hagen (1959) provides a comprehensive view of success and its relation to ability test scores. They studied 10,000 men who had been given a battery of aptitude tests in 1943 when they were applicants for Army Air Force cadet training. The battery included tests of general intellectual ability, numerical fluency, mechanical principles, and psychomotor coordination. They were followed up in 1955 and 1956 to determine their educational and vocational history. There were seven criterion scores: income, number supervised, self-rated success, job satisfaction, vertical mobility, lateral mobility, and length of time in occupation. The men were grouped into nearly 100 occupational groups. In general, out of all their correlations between tests and success within the occupations, only a slightly higher percentage were significant at the .05 level than would be expected by chance. They conclude: "This would suggest that we should view the long-range prediction of occupational success by aptitude tests with a good deal of skepticism and take a very restrained view as to how much can be accomplished in this direction" (p. 50).

Reilly and Chao (1982) reviewed the validity of possible alternative procedures that could be used instead of employment tests in selecting employees. Their extensive review examined various procedures including interviews, reference checks, projective tests, and academic achievement. Criteria for successful occupational performance varied. Their summary of the validity of academic performance variables is as follows:

> Average validities were computed against criteria of supervisory ratings and compensation. Eight validity coefficients with supervisory ratings yielded an average r of .14 (total $N = 994$). Nine validity coefficients with adjusted compensation yielded an average r of .27 (total $N = 1583$). Difficulties in interpreting these relationships have been noted above. Over all criteria 20 coefficients yielded an

> average r of .20 (total $N = 2272$). A recent independent review by O'Leary (1980) estimated an average validity of .17 (31 coefficients, total $N = 6782$). (p. 42)

They conclude:

> Results of a large number of studies investigating the relationship between academic performance and success in various occupations are less than impressive. Nelson (1975) noted that, ". . . a simple and direct application of grade point average, class standing, or similar academic achievement measure has little merit in valid job-related selection systems." . . . overall there is insufficient support for viewing grades or rank-in-class as a promising alternative. (pp. 43–44)

In general then, the consensus of previous reviews is that academic ability and grades are only slightly related to adult accomplishment. However, most of the reviews have been essentially summaries across studies to obtain an average correlation. The present review attempts to examine the evidence in greater detail by (1) carefully examining the criteria used in the studies, (2) looking at the special factors involved in different situations, e.g., medical practice, (3) using broad definitions of "success" as well as specific criteria, (4) looking at the evidence from high school and college samples, and (5) attempting to ascertain the underlying relationship between academic ability and accomplishment.

METHODOLOGY

Because the area of high-level accomplishment is so broad, extensive efforts were required to uncover the materials eventually reviewed. First an ERIC system search was made for studies relating accomplishment, achievement, or creativity with grades, academic ability, or test scores. Then a systematic search was made through *Psychological Abstracts* and *College Student Personnel Abstracts* on the same topics. Finally a systematic review of every issue of 19 journals was conducted for the years 1966 through 1984. These journals were:

American Educational Research Journal
American Journal of Sociology
American Sociological Review
Applied Psychological Measurement
College and University
Educational and Psychological Measurement
Intelligence
Journal of Applied Psychology
Journal of College Student Personnel
Journal of Counseling Psychology
Journal of Creative Behavior

Journal of Educational Measurement
Journal of Educational Psychology
Journal of Human Resources
Journal of Vocational Behavior
Measurement and Evaluation in Guidance
Personnel Psychology
Research in Higher Education
Sociology of Education

In addition, the research reports of organizations such as Educational Testing Service (ETS), the American College Testing Program, and the American Council on Education were reviewed for research results relevant to the topic. Finally, the books dealing with creativity, accomplishment, success, and human performance in the ETS, Rider College, and Princeton University libraries were examined. When any article or report of research was uncovered, the references to other literature were also examined. Any other reference that also seemed to deal with the topic was looked up. Thus, eventually, a reasonably thorough examination of the available literature was completed. However, in many cases this was just the beginning. The results of interest were often buried in obscure, hard-to-find technical reports or were hidden away in appendixes. Since many of the articles were not mainly concerned with the question of the relationship between academic talent and accomplishment, it was necessary to review carefully a wide range of articles that appeared potentially relevant to the topic.

Only the studies which included some real-life accomplishment or creative behavior were included in this review. Studies which concentrated on "creative personality profiles" and other data which do not demonstrate *actual achievement* were excluded. Studies of children were also excluded, since few children are capable of attainments of general social value. Although studies of the personality characteristics associated with accomplishment are important (e.g., Dellas and Gaier, 1970; Golan, 1963), the personality traits of achieving individuals will be discussed only when they shed light on the question of the relationships of real-life accomplishment and academic ability.

The subject of this review is related to the studies of the relationship of creativity and intelligence. It differs from those studies in its concentration on real-life accomplishment. Most of the creativity-intelligence studies have only examined the correlation between tests of creativity and tests of intelligence. Unfortunately, few "creativity tests" have been validated against real-life criteria of creative accomplishment, and when they have, they have done poorly (Baird, 1972a, 1972b; Crockenberg, 1972). There is a large difference between a child's ability to think of 20 uses for a brick and the publication in a scientific journal of an article describing the results of research.

Different measures of academic ability and success were used in the stud-

ies. In most cases, the measure was an academic admission test or grades received in an academic institution. In other cases, the measure used was a high-level verbal aptitude test, and in still others, an intelligence test. The latter were included because most intelligence tests are ultimately validated against grades or some other form of academic success.

As a recent review of over 300 references of intelligence testing (Joseph, 1977) concludes:

> It appears that the present testing techniques employed in test construction and methodology in the United States have been derived from and given impetus by Binet's original scale of 1905. This scale and his others which followed (1908, 1911) were a reflection of: school related abilities and not to be used to get at any congenital or acquired determination of the deficiencies reflected in the test results (Binet, 1980; Goodenough, 1969; Freeman, 1955; Edwards, 1971). The testing movement in America to follow (Goddard, 1910; Terman, 1961; Yerkes, 1915; Wechsler, 1939; Performance Testing, 1917; Group Testing, 1917; etc.) all seem to have built more or less on the basic Binet model (both theoretically and methodologically) and thus the tests as derived and validated against the original and revised Binet scales, did then, and still do, reflect what the Binet scales reflected: higher mental processes presumed to comprise intelligence as it is reflected in a school environment (Binet, 1980; Goodenough, 1949; Edwards, 1971; Zach, 1972). (pp. 80–81)

Making much the same point, Anastasi (1976) has written:

> Typical intelligence tests designed for use in our culture with school age children or adults measure largely verbal abilities; to a lesser degree they also cover abilities to deal with numerical and other abstract symbols. These are abilities that predominate in school learning. Most intelligence tests can therefore be regarded as measures of scholastic aptitude. (p. 350)

Thus, some studies which relate intelligence scores to attainment will be reviewed along with those using tests that are clearly measures of academic ability.

The most difficult problem in this area is to define the criteria of "success." A great variety of criteria have been used with many different definitions, even within the same category. For example, the criterion of scientific publications has been studied in a number of ways and is the subject of a small literature. The measures used have included self-reports of publications, counts from vita, weighted counts, counts in various journals, citations to articles, and indices of citations per article. However, as diverse as these criteria are, they can be arranged in a general order of significance and clarity as indicators of "success" and the research related to them similarly arranged.

The plan of this review is to discuss the criteria and research in an order

ranging from those most clearly representing high-level achievement to those representing general "success." First to be discussed will be the studies of the highly creative conducted by the Institute for Personality Assessment and Research at Berkeley. Next, the work of scientists and physicians will be considered. The literature on success in the upper levels of business and industry will then be examined. Then we shall turn to the studies of the creative and significant accomplishments of college and high school students. Studies of general "success," including sociological studies of occupational attainment, will be examined in the following section. Finally, the Terman studies of the lives of individuals with very high Stanford-Binet IQ's will be examined. Then the implications of research, along with research into the nature of human abilities and attainment, will be discussed.

STUDIES OF HIGHLY CREATIVE INDIVIDUALS

A group of studies which are directly related to the question of the relationship of academic ability to high-level accomplishment are those concerning highly creative individuals. Several research programs have been devoted to this question, especially the Institute for Personality Assessment and Research at the University of California at Berkeley, which conducted a series of studies of highly creative individuals in the fields of architecture, mathematics, scientific research, and writing. The samples, and the criteria used to select these groups, have been described by Barron (1965, 1969). The criterion was usually peer ratings based on the individual's accomplishments. In architecture, for example, an initial list of names was constructed from the names nominated as the 40 most creative architects in the United States by five senior professors of architecture at Berkeley. The individuals selected were also rated by 11 editors of major architectual journals. Finally, the architects who participated in the sample ranked each other. A comparison group of "representative architects" was also selected from listings in the *Directory of Architects*. This group matched the nominees in age and geographical area of practice. All the names of both the creative and representative architects were cast onto a single list, which was sent to 19 professors of architecture throughout the country, the original group of 5 professors, and 6 editors of architectural journals. All the names were ranked on a seven-point scale in terms of creativity. The "creative" group received significantly and markedly higher ratings. The latter were invited to the Institute at Berkeley, where they were subject to intense assessment, as were the creative members of the samples in other fields.

The importance of these studies is that the definition of creativity and accomplishment was carefully constructed, validated, and set at a high level. Only those who had made truly significant contributions to a field were

defined as creative; those who were merely productive were excluded. By calling on expert judgment within each field, the groups identified almost certainly included the most important figures in the fields. This definition would probably satisfy even the most skeptical within those fields.

In some cases "representative" groups were also invited to the live-in assessment, but in most cases they were not. Most of the data on the representative groups were collected by mail. The measure of intelligence used in the original studies was the Terman Concept Mastery Test, a very difficult test of vocabulary and analogies.

> Because of security regulations governing the use of the Concept Mastery Test, it could not be administered either to the comparison group of writers or to the comparison groups of architects. The only sample for which a true comparison group is available is the Creative Women Mathematician sample, and the observed difference between the "creatives" and the "representatives" among women mathematicians favors the former and is statistically significant. (Barron, 1965, p. 69)

However, within the creative architects, MacKinnon (1962) found that the correlation between Concept Mastery Test scores and rated creativity was −.08, and Gough (1961) found a similar correlation of −.07 among the scientists, results that might be expected, given the narrow range of ratings within the creative groups.

In a description of later research, MacKinnon (1968) reported that

> we have returned subsequently to our architects, mathematicians, and researcher scientists and administered to as many as were willing to cooperate the Wechsler Adult Intelligence Scale.
>
> We have divided the samples of architects and research scientists into three subsamples, ranging from the most creative to the least creative. Each sample of mathematicians, one male and one female, has been divided into two groups, a creative group and a comparison group. At this point, our most striking finding is the lack of any significant difference in IQ among the subsamples characterized by different levels of creativeness. The mean IQs for the three groups of architects are 132, 131, and 130; for the research scientists, 132, 132, and 132; for the male mathematicians, 135 and 133; and for the female mathematicians, 129 and 133. The ranges of IQs are similarly comparable from subsample to subsample: for architects, 120–145, 117–142, and 119–143; for research scientists, 120–141, 121–142, and 114–142; for male mathematicians, 118–152 and 126–138; and for female mathematicians, 118–140 and 118–145. (p. 107)

MacKinnon and Hall (1972) have reported these results in more detail along with the results of multiple regression analyses using rated creativity as the criterion, and concluded that

> scoring as more intelligent than a colleague does not guarantee that one will surpass him in creativeness, as data which there is not time to present in detail

> convincingly demonstrate: in every group the multiple regression equation to predict creativity from WAIS scores failed to approach significance in cross-validation. In contrast, multiple regression solutions to predict the creativity of our subjects from the scales of the Strong Vocational Interest Blank, the Study of Values, the California Psychological Inventory, the Myers-Briggs Type Indicator, FIRO-B, and the Gough Adjective Check List all cross-validated at the .01 level of significance or better (Hall and MacKinnon, 1969).
>
> Above a given minimal level of intelligence required for the successful practice of one's profession, which in the groups we have studied is quite high, what is most importantly determinative of creative performance is not a higher level of intelligence per se but particular constellations of non-intellective traits. They are the factors that make the difference between a successful practitioner of a profession and one who practices it creatively. (p. 520)

Thus, the Berkeley studies generally showed that within highly creative professions there are no consistent differences in accomplishment related to intelligence. As we shall see in the National Merit and Terman results, among groups of academically highly able individuals, the differences in accomplishment seemed due to variables other than intelligence. However, it should be reemphasized that the groups studied at Berkeley are very bright on the average. The typical IQ of 132 places these groups in the top 2% of the adult population. Even the typical lowest IQ of 119 is at the 89th percentile. Clearly, to enter their professions, these people had to have very high academic ability.

A few other studies have also examined highly creative individuals and compared them with their peers. For example, Cross, Cattell, and Butcher (1967) compared Sixteen Personality Factor Questionnaire (16PF) responses of 63 artists selected for having given clear evidence of unusual talent in drawing or painting with 63 controls who had approximately equal educations and worked in similar settings. They found many personality differences, but no difference on the 16 PF measure of "intelligence," although both groups scored quite high on this measure. The creative artists showed more dominance, self-sufficiency, and bohemianism, and less ego strength, self-control, discipline, and superego.

In an earlier study, Cattell and Drevdahl (1955) asked a panel of experts to choose three groups from rosters of their professional society members. The groups were eminent researchers, eminent teachers, and eminent administrators. As expected, the three groups differed from the general population in a number of predictable ways, including having a very high group score on "intelligence." Cattell and Drevdahl also compared the three groups. Researchers were more concerned with their internal thoughts (schizothymia) and more self-sufficient than either teachers or administrators. They also had less ego strength, more radical attitudes, and more "bohemian con-

cern." However, researchers were not more "intelligent" than administrators or teachers. Of course, we should not necessarily assume that administrators or teachers are any less creative or show any lower level of achievement just because their accomplishments are in an area more difficult to evaluate. In any case, this study does not show any difference in the intelligence of the criterion groups.

In sum, comparisons of the intelligence of highly creative professionals with their peers reveal few differences. Of course, one might not expect any large differences within such highly selective groups. However, there are consistently large differences between the creative groups and the general population in measures of intelligence. That their less creative peers were equally bright suggests that a certain level of ability is needed to enter certain fields, but that ability may not discriminate within the fields. In contrast, all of these investigations found fairly large and consistent differences in the personalities and values of the creative professionals and their less creative peers. These have been summarized by Barron (1965) and MacKinnon and Hall (1972), and include a strong need for independence and autonomy; a high acceptance of impulse; high personal dominance; rejection of conformity pressures in thinking; a detached attitude in interpersonal relations; risk taking; and a liking for order and method combined with a fascination with disorder and exceptions. That these kinds of differences appear within such highly self- and educationally selected individuals suggests that, beyond the level of ability needed to qualify for various fields, other characteristics are needed to attain high levels of accomplishment.

This idea is related to the proposition of various investigators that there is a "threshold" effect in the relationship between intelligence and creativity. That is, creativity and intelligence are thought to be related up to some "threshold" value, say, an IQ of 120, above which they are independent (Barron, 1965). This idea has been criticized by McNemar (1964) and others, and research on the topic has not tended to support it. However, it leads to another conceptualization, that of a "fan-shaped" distribution between creativity and intelligence where "at the high IQ levels there will be a very wide range of creativity, whereas as we go down to average IQ, and on down to lower levels, the scatter for creativity will be less and less" (McNemar, 1964, p. 879). We shall return to this idea later.

HIGH-LEVEL PROFESSIONALS

The studies described in the last section were based on very distinctive subgroups of high-level professionals, carefully selected for their unusual and significant contributions. One cannot generalize from these results to all professional work without great caution. What is needed is more informa-

tion about the relationship between academic ability and accomplishment among more typical high- and middle-level professionals. This section reviews two types of studies of the high and middle ranges: those using objective criteria and those using ratings. Medicine will be treated separately as a third group of studies.

Studies Using Objective Criteria

Probably the clearest criterion in scientific and technical performance is publication activity. Although there have been some controversies about the technical manner in which publications should be used in research studies (e.g., Clemente and Sturgis, 1974; Cole and Cole, 1967; Drew and Karpf, 1975; Porter and Wolfe, 1975), there is little controversy about the basic significance of publication activity. The scientific or scholarly journal article is the primary avenue of reporting scientific research. Furthermore, since journals exercise editorial control over what they print, a published article typically represents a level of scientific competence, and frequently a contribution to the field. Because of this, many faculty members and professional scientists are evaluated on the basis of the number of their publications.

One of the earliest investigations of predictors of scientific publications was part of an evaluation of the success of procedures used to select Veterans Administration (VA) trainees in psychology. In 1957 Kelley and Goldberg (1959) followed up two samples of graduate psychology students who had been VA trainees in psychology at the University of Michigan in 1947 and 1948. A wide variety of test and rating information had been collected on them when they were graduate students. Scholarly productivity, defined as number of listings of publications in *Psychological Abstracts,* was predicted in the 1947 sample by the Strong Vocational Interest Blank (SVIB) Psychologist scale (.33), Banker scale (−.29), and several other SVIB scales, and by a high-level verbal reasoning test, the Miller Analogies Test (.18). However, not a single variable correlating with scholarly productivity in the 1947 sample was found to correlate in the 1948 sample. In fact, there were no significant correlates of productivity in the 1948 sample.

The next major study was also an attempt to evaluate the success of selection procedures for a special program. The National Science Foundation (NSF) began its Graduate Fellowship Program in 1952. The program was designed to support able students in their graduate studies in one of the sciences. Under the direction of Calvin Taylor, followed by Lindsey Harmon, a continuous research program was designed to improve the procedures for selecting NSF Fellows. Graduate Record Examinations (GRE) scores and college grades were part of the selection information. Two groups of fellowship applicants were followed up intensively: the students who had

applied in 1955 and in 1956, when they were college seniors or graduate students with one or more years of graduate study. Criteria studied in follow-ups included Ph.D. attainment, income, number of publications or patents, number of times the applicant's works were cited in the literature, and several ratings derived from confidential reports made by the applicant's colleagues, subordinates, or former professors. Ratings of the applicant's overall performance from at least three people were sought in 1965, although in a few cases only one rating was obtained. The various criteria were treated with sophistication. For example, Creager (1963) developed a method to place the applicants in stanine groups, based on a coded index of the applicants' later publications and patents. The stanine system was important because of the skewed distribution of the production of articles and patents, a problem noted above.

The various selection variables were correlated with the criteria. Within-field correlations seem the most appropriate, since fields which differ on some predictor variables may also differ on the criteria. For example, chemistry students score high on the GRE-mathematics examination, and chemistry is a field with a high average publication rate. When all applicants were combined, there might appear to be a relation between GRE-mathematics scores and publication, simply because fields with different publication patterns were combined. In the fields where sufficient numbers of cases were available, the analyses were also conducted separately by year.

The results, as reported by Creager and Harmon (1966), are shown in Table 1. The GRE-verbal test was not related to income. It was related to the productivity index in three of the seven fields, although in two of these, the relationship did not hold from one year to the next. GRE-verbal was related to the citation index for two years in biology, one year in physics, and for one year in chemistry. It was also related to the ratings in biology, engineering, and psychology, and for one year in physics.

The GRE-quantitative test was related to income in mathematics and biology applicants. It was related to productivity among the 1956 biologists, the geologists, and the 1955 physicists. It was consistently related to the citation index in biology, chemistry, and psychology; it was inconsistently related in physics. It was related to overall rating in biology, engineering, geology, and, inconsistently, in physics.

College grade-point average was not related to any of the criteria except the overall rating in biology, chemistry, engineering, and, in the 1956 group, physics, and to income in mathematics. This result might be expected because of the very narrow range in grades.

The GRE-advanced tests are discussed last because they are, strictly speaking, less measures of general academic aptitude than measures of detailed understanding and mastery of each academic field. They are much

TABLE 1. Correlations of Income, Productivity, and Citations with Four Predictor Variables

	N	*Verbal*				*Quantitative*				*Advanced*				*Grade Point Average*			
		IN	*PR*	*CI*	*OA*	*IN*	*PR*	*CI*	*OA*	*IN*	*PR*	*CI*	*OA*	*IN*	*PR*	*CI*	*OA*
Biology																	
1955	182	06	08	*25*	*21*	*23*	14	*29*	*28*	08	*18*	*19*	*17*	-05	08	12	*31*
1956	160	11	*22*	*26*	*16*	*20*	*20*	*24*	*21*	*17*	*19*	*28*	*21*	04	-01	07	*21*
Chemistry																	
1955	185	-10	12	*30*	03	05	12	*25*	09	*20*	*27*	*28*	*18*	-06	05	13	*24*
1956	171	03	07	05	-02	-01	10	*17*	12	11	*43*	*35*	*21*	-07	04	11	*25*
Engineering	249	05	08	09	*17*	06	04	02	*23*	03	*17*	*13*	*27*	10	08	04	*18*
Geology	109	14	14	11	13	03	*21*	15	*19*	11	*33*	08	*42*	00	03	07	17
Mathematics	134	16	*26*	16	*17*	*29*	13	09	04	*17*	*34*	*20*	*18*	*-24*	05	05	-03
Physics																	
1955	192	-05	*21*	*18*	*21*	07	*24*	*30*	*19*	09	*29*	*31*	09	00	*14*	*14*	06
1956	160	02	04	05	02	*17*	13	12	12	*26*	*26*	*18*	*24*	-06	10	01	*21*
Psychology	73	-12	01	-03	*-23*	-03	12	*23*	11	09	03	*33*	-21	-11	04	-13	04

Adapted from Creager and Harmon, 1966.
Decimal points have been omitted.
Coefficients significant at the 5% level are italicized; those significant at the 1% level are italicized and underlined.
Columns headed IN give correlations with Income.
Columns headed PR give correlations with Productivity Stanine.
Columns headed CI give correlations with Citation Counts (1964 *Index*).
Columns headed OA give correlations with Overall Average performance ratings.

more narrowly defined and ". . . are designed to measure mastery and comprehension of materials basic to graduate study in major fields . . . an attempt is made to survey the field and to draw material from widely differing curricula. . . . The Advanced tests emphasize the basic concepts and principles of their subjects and include questions that require reasoning, analysis, and decisions based on knowledge of these principles." (Graduate Record Examinations, 1969, p. 5). The GRE-advanced tests were inconsistently related to income in biology, chemistry, and physics. They were related to productivity in every field but psychology. They were related to the citation index in every field except geology. They were related to the overall rating in every group except psychology and the 1955 physics group.

In short, measures of academic aptitude were not significantly related to the income criterion in most instances. The best predictors of the other criteria were GRE-advanced test scores, followed by the GRE-quantitative scores. However, most of the significant correlations were moderate. Of course, the correlations are almost certainly attenuated because of the restriction of range in academic ability. However, the restriction in range did not seem to affect the correlations of the GRE-advanced field tests. (Of 40 correlations between the criteria and the GRE-advanced tests 28, or 70%, were significant.) Thus, general academic ability did not seem to be as highly related to the criteria as did knowledge of a specific field. We shall discuss this point more fully later.

Using a more complex criterion, W. A. Owens (1969) collected information on the 1964 accomplishments and performance of 931 engineering alumni who were originally administered a variety of instruments in 1955 when they were (in most cases) juniors. The subjects were enrolled in a wide variety of institutions across the country. The original instruments included a biographical information blank, interest measures, an application of mechanisms test, a power source apparatus test, and, for 457 members of the sample, the American Council on Education (ACE) Psychological Examination, a measure of academic ability. The criterion was a summed score based on professional papers, professional journal publications, development or improvement of products or processes, and, given most weight, patents held, pending, or disclosed. Although the ACE examination had no significant relation with this criterion, several of the other measures had significant, if small, relationships. In addition, self-reported "academic achievement"—consisting of ranking high in one's class, being a member of an honor society, being a scholarship winner, etc.—had a significant correlation of $-.19$ with the criterion.

In a small study within a very specialized area, Gertler and Meltzer (1970) developed an equation from a study of 47 Ph.D. alumni of the industrial relations program at Carnegie-Mellon University. They predicted the per-

formance of the Ph.D.s during their careers in research, using information available at the time applications were submitted. The researchers developed an index of research publications, adjusted for the quality of the publication, as their measure of performance. Undergraduate grades, age at time of application, and previous graduate training appeared to be important for the prediction. Scores on standardized tests (GRE and the Admission Test for Graduate Study in Business) did not discriminate within the range covered by the sample. They interpreted grades as a measure of the motivation to succeed.

Folger, Astin, and Bayer (1970) studied the largest sample of any reviewed here: 6,300 doctorate recipients (1957–1959) in mathematics/statistics, physics, chemistry, biochemistry, and psychology who responded to the 1964 National Register file of recent doctorates. The criterion measures were number of citations to each sample member's works in 1964 and 1965. By searching high school records, ability measures were located for many in the sample. The correlations between these measures and citation counts in the above fields were mathematics/statistics, .04; physics, .10; chemistry, .07; biochemistry, .04; and psychology, .07. The correlation was significant only in physics. Quality of graduate department and time taken to attain the degree were correlated with citations. This study was limited by the availability and comparability of ability tests, the fact that they had to be equated, the long interval between the testing and the criterion, and the relatively short period of professional life covered, all of which would attenuate the size of the correlations. However, it is based on a large sample, and uses the criterion of citations, which some researchers have recommended as the best single measure of scientific impact.

One possible explanation for the low correlations between academic ability and productivity measures is that specialties within professions will make different uses of their abilities. For example, Marston (1971) found 111 University of Southern California Ph.D.s in psychology who graduated between 1952 and 1966. Marston correlated their scores with their *Psychological Abstracts* count as measured by weighted mean number of publications per year. Combined GRE-verbal and GRE-quantitative scores correlated −.05 among clinical Ph.D.s and .18 among nonclinical Ph.D.s. Because of skewness in the distribution of the criteria, point-biserial correlations were also run. All were nonsignificant. Weitzman (1972) suggested that skewness may be because only the few very high scorers published. This possibility was investigated by Schrader and by Clark and Centra in studies to be described later. Creager, in the study described earlier, also analyzed his sample by type of employment—academic, industrial, and governmental. Although there was a slight tendency for the correlation in industry to be

lower, the overall patterns were very similar to those described before. Thus, there is little evidence that results are very different in subfields.

Two studies with similar drawbacks and advantages were conducted by Kaufman and Hansen and Nevjhar. Kaufman (1972) studied 110 engineers from three technical organizations. They were administered an engineering achievement test (similar to the Undergraduate Program Field Tests in Engineering) shortly after they obtained their college degrees. They were followed up approximately 14 years later; and data from the first eight years, middle three years, and last three years were analyzed. Criteria were claimed range of area of accomplishment (diversity), competency in those areas (competence), supervisory ratings (performance), number of publications, and number of patents.

Achievement test scores were related to the number of patents in all three periods with correlations of .29, .34, and .31 to papers produced in the first period (.19) and the third (.23) but were unrelated to any of the ratings of competence. In a reduced sample of 32 engineers, scores were related to claimed competence (.38).

In the second study, 115 students who were enrolled in the Science Honors Program (for high school students) at Columbia in 1959 took the Pre-Engineering Ability Test (PAT) in addition to the engineering achievement test. Hansen and Nevjhar (1973) found that the PAT predicted the number of publications reported in a follow-up 12 years later, with a correlation of .26 with the mathematics test on the PAT and .31 with the achievement test. It also predicted whether the students obtained advanced degrees.

Schrader (1978) conducted a study that carefully defined both the sample and the criteria. Schrader studied a sample of psychologists who had earned a doctorate in psychology in 1963–64, who had earned a bachelor's degree between 1954 and 1961, and who had retrievable test scores either on the SAT or on the aptitude test and an advanced test of the GRE. The final sample sizes were 128 for SAT scores and 155 for GRE scores.

Measures of attainment included citation counts obtained from the *Social Sciences Citation Index* (SSCI) and the *Annual Review of Psychology,* and publication counts obtained from *Psychological Abstracts,* all based on entries between 1972 and 1975. Other criteria included number of times the subject was listed as a first author, and election to fellow status in the American Psychological Association (APA). The rating of the graduate faculty in psychology for each psychologist's doctoral university as reported by Cartter (1966) was included in the study, along with various other biographical variables.

Since the distribution of SSCI citations was skewed, Schrader included a normalized citation index as well as raw number of citations. Schrader

found that although SAT-verbal scales were not significantly related to any of the criteria, SAT-mathematical scores were correlated with normalized SSCI citation counts (.18) and raw counts (.26). GRE-verbal and quantitative scores were significantly correlated with most of the criteria. The GRE-advanced test was also correlated with all of the criteria except attaining fellowship in the APA. The correlations for GRE-verbal, GRE-quantitative, and GRE-advanced, respectively, with each of the publication and citation criteria, were as follows: for *SSCI* citations (raw) .26, .28, and .40; for *SSCI* citations (normalized) .28, .19, and .45; for citations in the *Annual Review* .21, .30, and .32; for total *Psychological Abstracts* count .17, .28, and .32; and for publications as first author .15 (not significant), .26, and .33.

Schrader also examined the distribution of the number of publications and citations across score groups. He found that the highest score group had the highest number of publications and citations, but that the lowest group had the next highest number, and the middle group, the lowest. The pattern of relationships, with the advanced test having the best correlations with the criteria, is similar to the results of Creager and Harmon.

Finally, a recent study by Clark and Centra (1982) seems to provide the most comprehensive analyses of the personal and situational influences on productivity. Clark and Centra studied two samples of doctoral recipients. The first was a sample of alumni of Ph.D. programs in chemistry, history, and psychology programs (Clark, Hartnett, and Baird, 1976), who had received the doctorate between 1970 and 1972 and were followed up in 1975. The second consisted of men and women who had received doctorates in 1960 and in 1968 who were followed up in 1973 (Centra, 1974). The criterion was number of self-reported publications. To check on the accuracy of these reports, the authors compared the reports of the male alumni in psychology who participated in the first study with the number of their entries in *Psychological Abstracts*. The correlation was .84, which seems quite reasonable, since the correlation between the count for 1967–1975 with the count for 1967–1977 was .96. Furthermore, since *Psychological Abstracts* does not abstract all journals, it also seems reasonable to suppose that many in the sample had published in journals not included in that count. Thus, the self-report of number of publications appears fairly accurate. GRE-verbal and quantitative scores were found for the subjects in the first sample. Because there were so few women, they were excluded from the analysis. The resulting sample consisted of 239 chemists, 142 historians, and 221 psychologists, all of whom had at least one GRE score. In chemistry, the correlation of number of articles and book chapters with GRE-verbal was −.02; with GRE-quantitative it was −.01; and with GRE-advanced it was .15. For only those chemists engaged in research in business or government, they were .13,

.11, and .05. For all historians, these correlations were −.24, −.14, and .00. For all psychologists, the correlations were −.05, −.02, and .02.

Clark and Centra also examined the distribution of a number of publications by GRE scores. The distributions were essentially flat, with no particular trend. In fact, the largest number of publications was reported by the lowest scoring groups in all three fields.

In the second sample, GRE scores were found for 94 respondents in the social sciences, 115 in the biological sciences, and 103 in the physical sciences. To create more stable variables, certain information was combined. A "productivity" measure was constructed by weighing the number of articles published, the number of books as sole or senior author, and the number of books as junior author or editor. Similarly, an "academic ability" measure was constructed by weighing GRE-verbal, GRE-quantitative, and GRE-advanced scores. There were no significant relationships between "productivity" and "academic ability."

Clark and Centra also used the technique of path analysis in both samples to determine the influences on productivity and income. In both studies, for the purpose of the analysis, an "academic ability" and a "productivity" factor were derived. In the first study, in chemistry, academic ability had no relationships (no path coefficient) with any other variable, including productivity. In history, academic ability had a coefficient of −.27 with productivity and no other variable. In psychology, academic ability had a coefficient of −.21 with productivity and no other variable. In the second study, among physical scientists, academic ability was related only to the prestige rating of the department that awarded the degree to the respondent. Among biological scientists, academic ability was unrelated to all other variables. Among social scientists, academic ability was also related to the prestige of the department and had a coefficient of .26 with productivity.

The most consistent influence on productivity across the six samples studied was the nature of the current position: those who were working in positions that emphasized research were more productive than those in other positions. There were a few, chiefly indirect influences of the rated quality of the Ph.D. awarding program.

These analyses are important because they show the interconnections and structural influences of variables on productivity, and thus provide much more information than simple correlations, which may be due to other factors than the two variables being related. They show that, in these samples, academic ability had at best an inconsistent relationship with productivity. However, the samples in the first study had been working for only three years, and the average number of publications was low, so that only a few cases might have altered the correlations considerably. The samples in

the second study lumped together fields that may have very different publication patterns; e.g., psychology and history were both included in the social sciences, but the average number of journal articles is much higher in psychology than in history. Thus, any differences related to academic ability may not appear because the field differences mask them.

The studies reviewed here have studied a variety of samples, with differing time frames, and were conducted for a variety of purposes. However, several broad conclusions seem warranted. In general, the correlations between measures of general academic ability and the publications or citations were low and inconsistent.

The result that GRE-advanced tests tended to have the highest correlations with the criteria was found in the Creagar and Harmon study and one of the Schrader studies, and in some of the Clark and Centra samples. The pattern in those studies suggests a general possibility that will appear in other studies which will be reviewed in later sections; the closer in time or more similar the test to the criterion, the higher will be the correlation. The GRE-advanced tests measure knowledge in a specific field, rather than general abilities. Thus, they represent measures of the academic preparation of individuals and, possibly, their motivation to learn and their interest in a field, all qualities that presumably would be related to high-level professional behavior.

Ratings of the Performance of Scientists and Technicians

A second group of studies is concerned with the general professional performance of scientists and technicians, as assessed by their superiors or peers. Although ratings have various problems (see, e.g., the discussions by Anastasi, 1976; McCormick and Tiffin, 1980), they also have several advantages. For example, Anastasi writes of ratings as

> an evaluation of the individual by the rater on the basis of cumulative, uncontrolled observations of daily life. Such ratings differ from naturalistic observations in that the data are accumulated casually and informally; they also involve interpretation and judgment, rather than simple recording of observations. In contrast to both naturalistic observation and interviews, however, they typically cover a longer observation period and the information is obtained under more realistic conditions. (p. 609)

Ratings may be especially appropriate for scientists and technicians, since they often produce products in the form of experiments, studies, reports, inventions, or improvements that can be used as one basis for their evaluation. Thus, the ratings may be made on a more objective basis than in some other fields of activity. These ratings have slowly increased in sophistication.

For example, in a comprehensive study of engineers conducted at Educational Testing Service, great effort went into developing and refining the rating scales, but chief reliance was placed on an overall rating.

Hemphill (1963) studied 448 newly hired engineers in five companies whose performance was rated by their supervisors after they had been on the job for two years. In the total sample, overall performance ratings were correlated .19 with a verbal reasoning test, although they were not correlated with three other tests. The correlation of verbal reasoning varied by type of work; it was not significant with ratings of those who were developing and utilizing personnel, but it was .32 with those who persuaded and negotiated with others. In two other samples of experienced engineers, performance ratings were correlated with a numerical relations test in two of seven areas of specialization.

In a study by Jones (1964), 25 managers rated 88 industrial scientists and technologists in a large company, using a weighted creativity rating scale (Sprecher, 1954). (It correlated .88 with a simple global rating of creativity.) A logical reasoning test ($r = .31$) and a mathematical reasoning test ($r = .29$) were related to the criterion as was an ideational fluency test ($r = .33$).

Kaufman (1972), in a study reviewed in the last section, included supervisors' ratings of the performance of engineers among several other criteria. Achievement test scores were unrelated to this criterion at any stage of the engineers' careers over 14 years.

Gough (1976) administered a battery of tests to 45 professional research scientists, who were also rated on creativity by an average of eight peers and two supervisors. Then he standardized and summed the ratings. Their reliability was .77. In another sample, senior honors engineering students at Berkeley were rated by their professors using the same criteria. Among the scientists, neither the Minnesota Engineering Analogies Test nor the Concept Mastery Test were correlated with the criterion, but a "scientific-word association test" was. (This test presented a scientific word, such as "neutron," and asked the subject to indicate the first word that came to mind.) In the student sample, the ability tests correlated with professors' ratings .34 and .33, respectively.

An increased level of sophistication is shown in a study by South (1974). For 130 young engineers, South used rating scales developed by factor analysis. These scales were correlated with a large number of tests. Various academic ability and intelligence tests were positively related to communication skill (the correlations ranged from .22 to .30), and technical knowledge (.25 and .31), but negatively related to administrative ability ($-.22$ to $-.30$), and "motivation" ($-.24$ to $-.32$).

A more comprehensive approach was used in two studies by Muchinsky and Hoyt. Muchinsky and Hoyt (1974) followed up engineers who had been

freshmen in 1956, 1957, or 1958 and graduated from the College of Engineering at Kansas State University. All had been out of college 5 to 10 years. Supervisors of 127 of the sample rated them on 10 traits and gave them an overall rating, and a weighted rating was devised. In addition, a rating of their best vocational achievement was made by faculty members. The subjects provided data on their salary and two self-ratings. The ACE-quantitative scores correlated with 3 of the 15 criteria: .26 with the weighted overall rating, .37 with rated written communication, and .40 with rated persuasiveness. The ACE-linguistic score was unrelated to all 17 criteria.

In an earlier study, Muchinsky and Hoyt (1973) used the same sample of engineers and the same criteria but used overall grade-point average (GPA), senior GPA, "core GPA," and "design GPA" as predictors. Although overall GPA was related to only one of 15 criteria, a rating of creativity or originality ($r = .21$), senior GPA was related to ratings of creativity or originality ($r = .30$), the achievement rating by faculty ($r = .28$) and overall occupational rating ($r = .23$); core GPA was related to ratings of precision and care ($r = .17$); and design GPA to salary ($r = .26$).

The most careful criterion in the studies reviewed here was developed by Andrews (1975), who examined the relationship between verbal ability and the quality and quantity of scientific output. Data were obtained from 115 sociologists, psychologists, and medical doctors who had directed research projects on the social psychology of disease. Andrews obtained copies of the reports or major publications these scientists had identified as the most important they had written about their projects. These were abstracted and independently rated by one to seven members of the American Sociological Association (median of 4.5 raters per project). They were rated on: (1) innovativeness, or the degree to which the projects advanced new lines of research or theory; and (2) productivity, or the extent to which the projects add to knowledge along established lines of research or theory. These ratings were correlated with the scientists' verbal scores on the General Aptitude Test Battery (GATB). Productivity correlated $-.01$ and innovativeness $-.09$. Andrews did not find any special situations (e.g., among less experienced vs. more experienced, those in different supervisory roles, etc.) in which these correlations were altered. (Andrews does not present the average GATB scores or their standard deviation, so it is hard to estimate the ability level of this group.)

In summary, the evidence on the relationship between academic talent and ratings of engineers and scientists was mixed. Some of the studies found some relationship, usually small, but a few did not. It is striking, however, that four of these studies (Mednick, Gough, Andrews, and Taylor and Ellison) reported positive results for some type of assessment of creativity.

Perhaps, again, when a certain level of ability is reached, factors other than academic talent become more important for accomplishment.

Studies in Medicine

The prediction of success in medicine is treated separately for several reasons. Medicine is a unique profession, combining science with practice and technical knowledge with personal, even intimate, contact with patients. It involves complex professional roles, so it represents a challenge to the researcher. Finally, it has been and continues to be the subject of many studies. The relationship between medical education and professional duties is being examined in a continuous program of research by the Association of American Medical Colleges (AAMC), as well as by several other groups. The results of their investigations are reported in a specialized journal, *The Journal of Medical Education,* and in *The Proceedings of the Annual Conference on Research in Medical Education* sponsored by the AAMC.

Various studies in this literature have been reviewed by Gough (1967), Wingard and Williamson (1973), and Cuca, Sakakeeny, and Johnson (1976). One of the best studies in this literature was conducted by Peterson et al. (1956). The medical expertise of 88 physicians in general practice in North Carolina was rated by internists who observed their behavior in their daily office work. The physicians were rated on six dimensions of professional competence. None of the ratings was significantly related to their Medical College Admission Test (MCAT) scores.

Howell (1966) contrasted 156 United States Public Health Service physicians who were rated high on official (open-ended) efficiency reports with 156 who were rated low. The physicians were employed in a wide variety of settings. There were many significant differences on various personality tests, including Adjective Checklist scales, California Psychological Inventory scales, and the K scale of the Minnesota Multiphasic Personality Inventory, but none on two tests of values and several other tests, including the MCAT and the Public Health Service Professional Examination in medicine.

In a later study, Howell and Vincent (1967) studied the relationship between MCAT scores and annual supervisory ratings and an achievement examination measuring academic knowledge of medicine. The correlations between the MCAT scores and the ratings ranged from $-.05$ to $-.25$. MCAT scores were related to medical knowledge test scores with correlations ranging from $-.05$ to .62.

A number of studies of practicing physicians have been conducted by a research group at the University of Utah. The studies discussed immediately following, as well as later ones, were subsequently described in greater detail

by Price et al. (1973). This group developed 77 measures of on-the-job physician performance in three samples: 102 full-time faculty members of the College of Medicine at the University of Utah (Taylor et al., 1965); 190 certified Utah specialists (Richards et al., 1965); and 217 general practitioners (Price et al., 1964). Premedical GPA, GPA for the first two years of medical school, and GPA for the last two years of medical school were correlated with the criteria in the three samples. Only 3% of the correlations were significant at the 5% level, and more of these were negative than positive.

In a subsequent study (Jacobsen et al., 1965), the Utah group studied first a slightly different sample of medical school faculty members ($N = 61$). Undergraduate grades were significantly correlated with 5 of 25 criteria: the respondents' regular review of scientific literature (.29), public recognition for contribution (.32), cooperativeness in the research project (.40), achievement in education (.56), and negatively, teaching responsibilities (−.42). Medical school grades were significantly and positively correlated with academic orientation-teaching excellence (.35), participation in social organizations (.32), achievement in education (.51), and negatively correlated with academic seniority (−.46), and participation in professional societies (−.30).

In a second sample of 242 general practitioners, the average correlation across criteria was .02 for pre-medical school grades, .03 for grades in the first two years of medical school, and .05 for the last two years of medical school. However, these averages mask some important results. All three grade predictors were highly related to the "achievement in education" factor (.74, .97, and .95). Pre-medical school grades were positively correlated with youthfulness in getting degree (.27) and socioeconomic status of parents (.23), and negatively correlated with civic participation (−.24), keeping abreast of medical progress through courses and professional groups (−.30), and diagnostic thoroughness (−.21). Grades in the first two years of medical school were negatively related to recognition by hospital staff (−.25). Grades in the second two years correlated positively with civic participation (.20), "orthodox success image" (.26), and correlated negatively with youthfulness in getting a degree (−.24). Scores on the Medical College Admission Test were also correlated with the criteria in this sample. The verbal score positively predicted the size of the physician's practice (.38), and negatively predicted medical referring (−.22) and off-the-job socialization (−.24). The quantitative score was negatively related to orthodox success image (−.18), medical referring (−.27), and professional stability (−.31). The science score was related to the factors of group or clinic practice (−.21), hospital staff recognition (−.20), prolonged postgraduate training

(.18), off-the-job socialization (−.27), and professional socialization (+.21).

Altogether, this study suggests that complexity of the criteria in a single profession, and shows how measures of academic ability and academic success can have varied relationships to those criteria.

The last research report in the Utah studies (Price et al., 1973) summarized several additional studies, which used some combined samples and new samples. In the first of these later studies, the General Aptitude Test Battery (GATB) scores, grades, and Professional Aptitude Test scores of a sample of 31 medical students were related to their professional performance as physicians 19 years later. Overall only 5% of the predictive validity coefficients were significant at the 5% level of significance. However, this result may not be surprising, given the length of time between the testing and the gathering of the criterion data.

In another study, the various criteria were summarized into five performance criteria. In a combined sample of 333 physicians who had graduated from medical school on an average of 16 years earlier, premedical grades were not related to any of the criteria. Grades in the first two years of medical school were correlated with a summary score from the 80 criteria as weighted for discriminating superior physicians (.20), as were grades in the last two years (also .20). Grades in both medical school periods were correlated with judgments of the quality of a portfolio of each physician's history and accomplishments ($r = .33$ for first two years, .22 for last two years). Grades in both periods were related to a rating of the physician when his or her name was known to the rater ($r = .21$ for the first two years, .25 for the last two years). Grades in the last two years were also related to an "equally weighted" composite (.16). When the physicians were grouped by type of practice, grades did not significantly predict the criteria among general practitioners. Among specialists, grades in both medical school periods were related to all the criteria except a judgment of the quality of their contributions. The correlations ranged from .21 for the weighted composite to .41 with the rating when the physician's name was known to the rater. Grades were related to the judgment of quality of their contribution (.35) and the rating by name (.32).

In a second study with this sample, an additional eight output criteria were used. An "output composite" was correlated .21 with undergraduate grades and .23 with grades in the first two years of medical school. An index of "high relevance" was correlated .24 with grades in the first two years and .20 with grades in the second two years.

These correlations seem quite reasonable, even unexpectedly high when one considers: (1) the variety of the sample, which combined general practi-

tioners, specialists, and medical school faculty; (2) the average length of time—16 years—between the academic performance and the criterion data; (3) the combining of very divergent and detailed criteria, many of which would have no relevance to particular physicians, into total scores or ratings.

Wingard and Williamson (1973) reviewed 7 studies relating medical school grades to the performance of physicians and 20 related studies in other areas. Criteria ranged from ratings of the quality of their technique made by internists to elaborate factor scores. Their conclusion: "Although studies in this area are sparse, available research findings have demonstrated that little or no correlation exists between academic and professional performance" (p. 313). They also reviewed research on career performance in related fields and reached the same conclusion. They considered four possible explanations for the low correlation.

1. *Deficiences in present grading systems as not reflecting qualities needed in real-life work:*

> The role of the physician, whether as practitioner, investigator, teacher, or administrator, is basically that of the problem-solver: the physician must be sensitive to problems and be able to collect adequate data, conduct analyses, draw conclusions, communicate the findings, and organize human and technical resources to implement the solutions. Since, with few exceptions, grading does not attempt to utilize criteria of this type, it is likely that grading would be deficient in application. (pp. 313–314)

2. *The failure of selection procedures to include characteristics that are important in professional careers.* They point out that selection

> procedures may identify only those who are most likely to achieve success in a current educational program. Consequently, many students selected, often on the basis of Medical College Admission Test scores, have characteristics that may ultimately determine adequate performance, for example, professional integrity, concern for people, and the ability to relate and communicate interest in the concerns of the community served by the physician. The fact that such qualities are rarely weighted heavily in selecting students for medical training or included in the process of student evaluation might have a significant effect in distorting the relationship between academic and professional performance. (p. 314)

3. *Intervening experiences.* Physicians (and other professionals) hold internships, residencies, and fellowships, and obviously have many different kinds of career experiences. These career experiences may have as strong an effect as schooling.

4. *Grades as indicators of ability.* Some medical educators claim that grades only assess the potential of students' talents for career performance; achievement of this potential may not be predictable.

Thus, according to Wingard and Williamson, there are good reasons to

believe that medical school grades assess behaviors and performances that are different from those that are important in medical practice.

It might be argued that this lack of relation is not due to differences in the behavior or activities in academic and professional settings, but is due to the restriction of range in the academic ability of physicians and medical students. Since this argument may be put forward for many of the studies reviewed throughout this report, it seems reasonable to quote the response of Price et al. (1973) at length:

> Medical school grades are inadequate as guidance or predictive tools for later physician performance. Based on all available evidence, grade point average does not predict how well medical students will perform in medical practice. That is, regardless of any possible restriction in the range of talent sampled present academic grades do not differentially predict later performance.
>
> Medical school grades are inadequate as substitute criteria for on-the-job performance of physicians. Our research has shown that academic performance is independent of actual performance and typically comes out as a separate and independent factor. In other words, grades do not come close to being parallel forms to later criteria of professional performance. In fact, the correlations fall far short of being high enough for satisfactory reliability coefficients; but instead nearly all of them are so low as to question whether any of them were truly non-zero correlations. Thus, such measures are totally inadequate as either substitutes or early indicators of later performance. This conclusion would likely hold even if extremely generous correction for restriction of range of talent were applied, due to the consistently zero or low levels of correlation found between grades and actual performance measures.
>
> Correction for restriction of range in our data would yield a greater number of moderately high negative correlations than high positive ones, a troublesome finding, indeed, for school grades . . .
>
> Correction formulas for direct restriction of range (on grades, for example) are not highly corrective for near-zero correlations, especially in the case of multiple independent criteria. (Correction formulas for indirect restriction of range are, of course, even less corrective.) (pp. 15–17)

However, this argument still does not give enough attention to the fact that medical students are selected on the basis of academic ability and on the basis of a wide variety of other personal characteristics. That is, the academic ability of admitted medical students is so high that differences in their medical school grades and their subsequent performance may not be attributable to their ability. Furthermore, as research reported by Baird (1975) indicates, medical school students also tend to be relatively homogeneous on such characteristics as career values, self-conceptions, educational orientation, and family background. And, obviously, as premedical students they had very similar educational experiences prior to medical school. Thus,

medical students represent a group "restricted in range" in terms both of academic ability and variety of other characteristics. The argument also neglects some of the group's own findings that medical school grades did have some relationship with overall judgments of physicians' accomplishments (Price et al., 1973). That MCAT scores and medical school grades were not consistently related to narrower and more specific criteria of physician performance may have more to do with the complexity of the physician's role and the specificity of each situation rather than to the unimportance of academic ability in the physician's performance. Obviously, the basis for award of medical school grades could be improved to incorporate judgments of more characteristics needed in the physician's actual work, and selection decisions could place more weight on evidence of such characteristics in applicants. However, this still would not eliminate the importance of academic ability for successful completion of the medical school program and preparation for the work of the physician.

SUCCESS IN HIGH- AND MIDDLE-LEVEL MANAGEMENT

A large group of studies has been concerned with the prediction of success in management. Such volume might be expected, since companies and organizations naturally have a strong interest in locating variables that will help them select managers. For their practical purposes, the studies are most helpful. However, for the purposes of this review, these studies are less valuable. The definitions of managerial success vary from study to study; and some ostensibly objective criteria, such as salary, are much more problematical than they first appear. However, as an area that employs the largest number of college graduates, and one that is obviously realistic, managerial success is probably of more importance to more people than any of the other criteria examined in this review. The studies fall into four main groups: (1) those using some measure of salary as the criterion, (2) those using managerial level attained as the criterion, (3) those using ratings as the criterion, and (4) those using an overall index or composite as the criterion.

Studies Using Salary as the Criterion

This criterion has many advantages as an index of managerial success since salary is the ultimate indication of the value a company places on an individual. In that sense, it may be more indicative than job titles or managerial level. However, salary as a criterion must be viewed with caution. Obviously, if salaries at different times are compared, they should be adjusted for inflation. A high salary after five years in a company is more indicative of success for a person who began with a low salary than for one

who began with a high salary. The first person has made a great deal of progress; the second, relatively little. Salary schedules differ from industry to industry and company to company, so studies of salary conducted across companies need to be interpreted carefully. The studies reviewed in this section have dealt with these problems with varying techniques and with varying success.

For example, Harrell (1969, 1970) attempted to control for the type of the company in which Stanford MBAs were working. In the first study, three classes of Stanford MBAs who were working in big business firms were surveyed to ascertain their current salaries and incomes. The highest-earning third ($N = 55$) was compared with the lowest-earning third ($N = 55$) on a variety of measures administered during graduate study, including the Strong Vocational Interest Blank and the Ghiselli Self-Description Inventory. Second-year business school grades distinguished the groups, although undergraduate grades and Admission Test for Graduate Study in Business (ATGSB) scores did not. A variety of personality measures suggested the high earners were self-confident, ascendant, and had high energy. In the second study, Harrell (1970) compared the highest- and lowest-earning thirds who were working in small business. Again, neither undergraduate grades nor ATGSB scores distinguished the groups, although second-year business school grades did, as did an ascendance scale. Harrell (1972) repeated these procedures with the addition of two more classes and a time period up to 10 years, with essentially the same results.

Another strategy was used by Dodd, Wollowick, and McNamara (1970) who controlled for the level of education within one company by studying persons who had the same training for their position. They followed up 396 IBM maintenance technician trainees for 9 years. At the end of this period their positions ranged from low-grade technician (which was similar to their entry position), to high management positions. Salary was used as a surrogate variable for management success. Training grades and the Gordon Personal Profile "ascendancy" scale both correlated .23 with salary after 9 years, although the Otis intelligence scale did not.

Yet another strategy was used by Tenopyr (1969) who found that for 113 managerial personnel a verbal comprehension test was correlated ($r = .29$) with salary corrected for age and seniority. The Leadership Evaluation and Development Scale was correlated .36 with the same criteria.

For 136 alumni of the Carnegie-Mellon graduate management program, Weinstein and Srinivasau (1974) obtained salary data, adjusted for work experience. They related salary to predictor information. Grades in graduate management school were correlated .49 with salaries of those in line positions and .24 with salaries of those in staff positions. ATGSB scores, undergraduate GPA, and "scholastic recognition" did not survive a cross-validated

multiple regression analysis (these were the only data they reported) among either staff or line managers. Ratings of involvement in social and sports activities were also positively related.

Wise (1975) obtained data about the rate of salary increase among 976 college graduates sampled from a large manufacturing corporation. Background data were related to rate of salary increase. College GPAs were related to the rate of salary increase. College GPAs were related to the rate of salary increases in a regression (least squares) analysis along with such personal qualities as leadership, as indicated by college leadership activities. (Zero-order *r*'s showed the GPA was correlated .24, holding an M.A. degree .22, and leadership .26). Wise concludes:

> These findings lend support to the practice of selecting students on the basis of academic measures. But non-academic attributes, largely independent of academic characteristics, have also been shown to affect productivity. The two groups seem to be of approximately equal importance. In light of the use of the college degree as an occupational screening device, this suggests a second look at the practice of selecting persons for higher education solely or largely on the basis of academic aptitude or achievement. If persons were selected for higher education on the basis of their potential productivity in a chosen occupation, rather than their potential as future students, consideration of nonacademic as well as academic attributes would be necessary. (pp. 364–365)

Two studies suggest the importance of controlling for educational attainment. Kinloch and Perrucci (1969) studied a national sample of 143 organizations and nearly 4,000 engineers and managers. Of these, 1,142 subjects with 0 to 6 years experience were studied in detail. College grades had gamma coefficients of .42 with monthly salary, .38 with yearly salary, .18 with level of supervisory responsibility, .40 with level of technical responsibility, and .31 with participation in professional activities. Degree level, however, had gamma coefficients of .80, .62, .31, .60, and .57 with the same criteria, and prestige of college correlated .33, .23, .08, .25, and .23 with the same criteria. Unfortunately, this study did not attempt to control for the effect of obtaining advanced degrees and then examine the effect of grades.

Perrucci and Perrucci (1970) studied a sample of engineers who had received their B.S. degrees in engineering from the University of California at Berkeley and the University of California at Los Angeles from 1947 to 1961. A follow-up several years later gathered information about (1) their gross annual salary in 1961, (2) their average monthly salary in 1962, (3) the level of their technical responsibility, (4) the level of their supervisory responsibility, and (5) their involvement in professional activities. Gamma coefficients showed that college grades were related to each of the criteria as follows: (1) .40, (2) .35, (3) .17, (4) .27, and (5) .29. Degree level was even more strongly

associated, having the following gammas: (1) .75, (2) .57, (3) .21, (4) .58, and (5) .51. Again, no attempt was made to control for the effect of degree level in relating grades to the criteria, so that a more sophisticated estimate of the relationship between grades and success was not possible.

Finally, two studies illustrate the use of multiple salary criteria. Crooks and Campbell (1974) obtained data from 128 University of Michigan MBAs and 66 Cornell University MBAs six years after they had graduated from business school. A career history questionnaire was administered to obtain information about salary and salary progress, mobility since MBA, level of responsibility attained, and level in the management hierarchy. In addition, an executive position description questionnaire was administered. Predictor variables included undergraduate grades, ATGSB scores, business school grades, and business school faculty ratings on 13 scales plus an overall rating of performance. The authors used a variety of measures of current salary, increase in salary from starting salary, and rate of increase. Although business school grades were correlated .14 with three measures of current salary, undergraduate grades and ATGSB scores were unrelated. None of the academic predictors was related to the indices of increase or rate of increase in salary.

In another study, Pfeffer (1977) found that among 216 MBAs from "a large, prestigious state University" who responded to a follow-up, neither GPA nor ATGSB scores were related to salary (20 graduates a year from 1960 to 1974 were sampled), whether considering starting or current salary, or salary adjusted for inflation.

To summarize, most of the studies of salaries among managerial level employees found a low relationship between salary and measures of academic ability. (The same general results were obtained in studies reported earlier by Muchinsky and Hoyt, 1973, 1974.) In general, as noted earlier, the closer the content of the measure of academic aptitude or performance was to the actual duties of the current position, the higher was the relationship.

Studies Using Level of Management Attained as the Criterion

Like salary, the managerial level a person attains must be used cautiously as a criterion. Again, titles and levels differ from company to company. The same supposed level has entirely different meanings in different contexts. For example, some companies have only two or three vice presidents, each of whom has broad powers. Banks have traditionally had many vice presidents, many of whom must work within narrow areas. The issue is further complicated by the fact that the attainment of a high-level position represents a different level of accomplishment at different stages of the career. A vice presidency at 35 is usually a greater accomplishment than one at 55. Again,

the studies have controlled for these complexities with varying methods and success.

For example, Kraut (1969) conducted a study based on the notion that a high-level test is needed to discriminate among high-level managers, i.e., that the low correlation between academic ability and success in high-level accomplishment found in other studies is due to the easiness of the tests used, which would result in a narrow range of scores. Consequently, Kraut used the Concept Mastery Test and the Ship Destination Test, both very difficult tests. They were administered to 235 middle managers and 130 higher-level executives who attended advanced management training programs. The results showed that the number of position levels the managers had moved four to seven years later was unrelated to either test in either group. In explaining these results, Kraut argued that the number of managerial levels moved up is the best or most important marker of managerial success. He argued for a "threshold" effect: beyond a minimal level of talent needed to handle the work of management, no more is needed.

The Kinloch and Perrucci (1969) and Perrucci and Perrucci (1970) studies reviewed in the last section found gamma coefficients of .40 and .17 between college grades and the level of technical responsibility, and .18 and .27 between grades and level of supervisory responsibility. However, as noted in these studies there was no control for level of degrees attained.

Using the same five classes described earlier (Harrell, 1972), Harrell and Harrell (1973) compared the Stanford MBAs who had reached general management with those who were in marketing, finance, consulting, accounting, production, and engineering, including research and development. Second-year grades did not seem to be related to attainment of early general management positions. On the ATGSB, general managers had higher quantitative scores than those in marketing, but lower scores than those in production. There were no differences on ATGSB verbal or total score. Again, it is probably hard to distinguish among Stanford MBAs in terms of academic ability. General managers tended to be energetic, decisive, dominant, and extroverted.

Crooks and Campbell (1974), in the study described in the last section, used a variety of definitions of supervisory level attained and related them to ATGSB scores, undergraduate grades, and business school grades. ATGSB verbal and business grades were correlated .15 and .15, respectively, with a score reflecting long-range planning, and business school grades were correlated .17 with a score reflecting exercise of broad power. Some of the measures were correlated with whether the person administered an annual budget. Undergraduate grades were correlated $-.22$ with this criterion, ATGSB quantitative, $-.20$, and ATGSB total, $-.15$.

In sum, there is mixed evidence for a low relationship between supervisory level attained and measures of academic ability and grades.

Studies Using Ratings as Criteria

The advantages and disadvantages of ratings as criteria of success were discussed in the section on scientists. Although the advantages are similar, some of the disadvantages are greater in management. The behaviors that are being rated are not as clearly related to success and are often difficult to observe or rate. Again, studies have varied in the sophistication with which they have used ratings.

For example, Tenopyr (1969), in the study described in the last section, used only two ratings: supervisor's rating and a labor relations rating. Neither was significantly related to a verbal comprehension test. In a somewhat more complex study, Rowland and Scott (1968) used superiors' ratings of: (1) supervisors' characteristics, and (2) amount, and (3) quality of work done by their work groups. A measure of intelligence, the Purdue Adaptability Test, was unrelated to any of the criteria.

Pallett and Hoyt (1968) used a great variety of ratings. A sample of University of Iowa graduates who had graduated between 1954 and 1959 were followed up in 1964, or 5 to 10 years later. Those who were employed were rated by their immediate supervisors on rating scales which yielded 23 three-item scale scores of "elements of success in general business." In addition, overall ratings of "progress" and "potential" were obtained. These were correlated with scores on the Iowa College Scholarship and Placement Tests and with grades in the last two years of college. The former was available for 116, and the latter for 184. The scholastic aptitude test scores were related to five of the criteria: problem solving ability ($r = .20$), judgment (.20), accuracy (.27), dependability (.21), and written communication (.19); college GPA was not related to the criteria.

In studies reviewed in the section on ratings of scientists and engineers, Kaufman (1972) found a relation between supervisory ratings and measures of academic ability in one of the three samples studied. The Muchinsky and Hoyt (1973, 1974) studies found similar results.

Studies Using General or Combined Criteria of Success

Finally, a few studies have used general or composite criteria of success. These have varied. In a study of the interaction of various traits with motivation, Ghiselli (1968) judged 271 middle managers in a variety of businesses and organizations as "unsuccessful" or not. A measure of intelligence did

not correlate with this rating, whatever the motivational state.

The Standard Oil Company of New Jersey (SONJ) has assessed managerial success for many years. As reported by Laurent (1962), SONJ sought predictors of three criteria: relative position level attained, salary progress, and ratings of managerial effectiveness. These criteria were combined to form an overall success index, which was shown to be independent of age and experience. This was correlated with a variety of measures. The best correlates, in two samples, double cross-validated, one consisting of 222 managers and the other of 221 managers, were special biographical survey keys ($r = .63$ in one sample, .50 in the other), special Guilford-Zimmerman keys ($r = .31$ and .32), and a management judgment test ($r = .41$ and .47). The Miller Analogies Test correlated .18 and .17 with success, and a nonverbal reasoning test correlated .20 and .08. Unfortunately, the specific correlations in the biographical scale are confidential, so there is no hard information about the correlation of college grades with success. However, a personal communication from Laurent, as reported by Campbell et al. (1970), indicated that successful managers had been successful in college, were active in taking advantage of leadership opportunities, and were forceful, dominant, assertive, and confident.

A number of other industrial studies, such as those conducted at Prudential Insurance (Selover, 1962), have used academic ability measures, and the written reports suggest that they may be useful predictors. Unfortunately, the reports are frequently vague about specific results, which is often due to their desire for secrecy about their companies. However, Bentz (1967) studied the success of a wide variety of executives at Sears, Roebuck, and Co., and found that the highest median biserial correlations were, in descending order: Allport-Vernon-Lindsey political score (.28), Guilford-Martin self-confidence score (.25), Kuder persuasive (.21), ACE test total score (.21), ACE linguistic score (.21), and Guilford-Martin masculinity score (.21).

As Campbell et al. (1970) point out, most of these studies have technical weaknesses, such as lack of cross-validation, contaminated criteria, and inappropriate statistics. They are also difficult to summarize as a group because they have used different criteria, different predictors, and very different methods of assessing predictive accuracy. Finally, some of the investigations have been done on first-level supervisors instead of higher-level management officials. In addition, many of the studies are based on small samples, use poorly validated instruments, and demonstrate concurrent rather than predictive validity. However, the fact that tests of academic ability were correlated with the criteria provided some positive validity for such measures.

In a related publication, Dunnette (1971) described studies conducted at the firm of American Telephone and Telegraph that built on the earlier work

of Campbell et al. Eight behavior rating factors were developed (general effectiveness, administrative skills, interpersonal skills, etc., plus an overall staff prediction of the eventual success of the ratees). The company's test of mental ability was later correlated, along with the assessment ratings and other tests, with salary progress of college and noncollege men. Generally, the highest and most numerous correlations were with staff assessment judgments, group simulations of business dealings, and interview ratings. The highest correlations for the mental ability test and two measures of "success" were, for college men in Company A, .48 and .38, respectively; in Company C, .51 and .32; for noncollege men in Company B, .47 and .45; and in Company C, .52 and .28.

In a later AT&T study, Grant (1975) reported on the predictive power of an initial assessment in relation to management level reached eight years later. For 123 college men, the most important predictors were variables reflecting interpersonal skills, personal stability, administrative skills, energy, and ambition; scholastic ability correlated .19. Among noncollege men, the most important predictors were interpersonal skills and administrative skills. Scholastic ability correlated .31. There was no information about the ability level of the first group, except that they were college graduates selected by the company.

Because of the manner in which these last few studies have been reported, it is difficult to assess the adequacy of the samples, measures, or criteria. However, the Grant report suggests that academic ability has some influence on managerial success. This conclusion applies to all the studies in this section.

Summary

Korman (1968) examined a wide variety of studies published from 1947 to 1965 attempting to predict managerial performance. Criteria included ratings of performance, administrative level attained, salary, objective performance, and termination of employment. A variety of predictors were used, including verbal ability tests (Cooperative School and College Ability Tests, California Test of Mental Maturity, American Council on Education, Miller Analogies Test, etc.). Korman concluded that such tests had some value in predicting the performance of first-line supervisors but were less useful in predicting higher-level managerial performance. He argued that this was not because cognitive skills were unimportant at such levels but because the groups are so preselected that it would be hard to show a relationship. The later research reviewed here tends to corroborate that conclusion. The research also suggests that the measure of academic success most relevant to managerial work—business school grades—does have a

positive relation with managerial success. Again, the closer the content of the measure of academic ability was to the actual duties of a field, the better it predicted.

ACCOMPLISHMENT IN HIGH SCHOOL AND COLLEGE

To this point we have examined the relationship between academic ability or academic success and adult accomplishment. The criteria of accomplishment have varied, but all have had reasonable face validity. It could be argued that all represented some accomplishment or performance that is valuable in the real world of adult life. In this section we shall examine a large body of research concerned with accomplishment in the high school or college years. This level of accomplishment can vary considerably in its intrinsic importance and its relevance to accomplishment in the adult world. For example, an undergraduate who publishes an article in a scientific or scholary journal has met the same high-level standards faced by professional scientists or scholars. In contrast, a student who works on the school newspaper may only be fulfilling a requirement in a journalism class. The criteria used in most of the studies reviewed in this chapter were designed to cover a range of accomplishments from the private and fairly common (e.g., writing a poem for one's own pleasure) to the public and rare (e.g., winning a prize for a scientific experiment). These criteria are generally fairly similar to adult accomplishments, but are set at a somewhat lower level and within the school or college context. The behaviors have clear significance within those contexts, although their importance for the general society is not always entirely clear. However, they do represent accomplishments within a particular setting, and they are important as precursors of later attainments. Furthermore, a number of the studies of adult accomplishment that have been reviewed in earlier chapters have found that most people who achieve at a high level during their adult careers had also achieved in the same areas during high school or college.

Since the meaning of attainment is especially important in these studies, considerable attention will be devoted to descriptions of the development of criteria in the following pages. The studies fall into two categories: those that were conducted at the National Merit Scholarship Corporation and the American College Testing Program, and all others.

The National Merit Scholarship Corporation and the American College Testing Program Studies

The National Merit Scholarship Corporation (NMSC) was founded in 1955 with the purpose of identifying the nation's most talented high school

students and providing financial assistance for their college educations. Supported by funds from the Ford Foundation and the National Science Foundation, the NMSC tested several million high school students each year. After a number of studies of the predictors of the academic accomplishment of the very bright students who received scholarships, the NMSC research staff began to explore definitions of talent broader than that of academic ability. Since they were concerned about identifying students who would potentially make a creative contribution to society, as well as to identify those who were bright, the NMSC research staff began a series of investigations into the nature of creative accomplishments. Subsequently, the American College Testing Program (ACT) conducted a series of similar studies. These studies can be divided into the correlational studies, the distribution studies, and the technical studies. The correlational studies will be reviewed first.

In the first of the NMSC investigations, Holland (1961) reviewed the secondary school achievements of Merit Finalists and developed scales of "creative science performance" and "creative arts performance." Because all the scales used in subsequent studies follow the basic model Holland used in this study, his account of the scales deserves to be quoted in full:

> The criteria of creative performance were derived from a checklist of accomplishments assumed to require creative or original behavior. Creative performance is defined as a performance which is accorded public recognition through awards, prizes, or publication, and which may therefore be assumed to have exceptional cultural value. Because of the difficulty in arriving at a generally acceptable definition of 'creativity,' these criteria should perhaps be regarded as either 'notable scientific or artistic performance,' although we will refer to the criterion as 'creative' performance hereafter to enhance readability. With this definition as a guide, a list of 20 achievements at the high school level was derived by reviewing the secondary school achievements of Finalists from previous years. Items were divided by content into two scales: Creative Science (5 items) and Creative Arts (11 items). (Four of the original 20 items were omitted because they appeared to be inadequate signs of creative behavior.) (p. 137)

Holland found that the creative performance scales were basically unrelated to grades and academic ability. Furthermore he found that the scales and grades were correlated with very different measures. Many of the variables which had the highest correlations with the creative performance scales had negative or near-zero correlations with grades. However, this conclusion is based on results from an extremely narrow band of academic talent. The Merit Scholars were highly selected, not only in terms of academic talent—approximately the top 1% of applicants—but also in terms of their extracurricular activities, their reputation among school and local officials, etc.

Holland and Astin (1962) also found essentially no relation between college-level "creative" accomplishment and grades and academic ability in four separate samples of Merit Finalists in each year of college. In addition to the scales used in the Holland (1961) study, they developed a scale of social or leadership accomplishments. Holland and Astin studied the predictive validity of information collected before college over one, two, three, and four years. Again, no relation between grades or ability and creative and social accomplishment was found. They also found that grades and social and creative accomplishments had different patterns of correlations with the predictive variables, which included the 16 PF and California Psychological Inventory.

Nichols and Holland (1963) examined 154 predictors of the first-year college achievements of a sample of Merit Finalists in academic areas and in the areas of science, art, writing, dramatics, music, and leadership. Items similar to the ones used in this and subsequent studies are shown in Table 2. To study the possibility that different predictors could be related to all accomplishments and to rare accomplishments which involve public recognition, analyses were conducted both ways. Essentially no relationship between grades and accomplishments was found for the male sample. However, among females, there was a correlation of .30 with all science accomplishments, .32 with rare science accomplishments, and .36 with rare writing accomplishments.

A subsequent study by Holland and Nichols (1964) was distinguished by the cross-validation of its results and by the "potential" scales which the researchers developed, based on the results just described, to assess lower-level activities which might predict accomplishment. These scales were described as follows:

> To predict student achievement in artistic, musical, literary, scientific, dramatic, and social fields, six 'potential for achievement' scales were constructed for each sex. Students falling in the upper and lower 27% on checklists of accomplishments for these fields in high school were compared for their preferences for 273 daily activities, hobbies, reading habits, school subjects, sports, etc. Typical items included working on guns, building scientific equipment, playing chess, going to a public library, giving talks, collecting rocks, playing charades, and drawing cartoons. The 15 most discriminating items were selected for each of the six Potential Achievement Scales for each sex. (pp. 55–56)

These scales were developed to meet the problem that many achievements are quite rare; the rare achievement, however, was probably preceded by a variety of lower-level activities by which the students' talents and skills were developed.

Other predictors which had proven to be useful in earlier studies were also

TABLE 2. Examples of Items Used in High School and College Achievement Scales

Area	*High school items*	*College items*
Leadership	Organized a school political group or campaign Received an award or special recognition for leadership of any kind Was elected to one or more student offices	Active member of four or more student groups Served on a student-faculty committee or group Elected as one of the officers of a class (freshman, sophomore, etc.) in any year of school
Art	Exhibited a work of art at my school (painting, sculpture, etc.) Had photographs, drawings or other art work published in a public newspaper or magazine Won a prize or award in a statewide or regional artistic competition (sculpture, painting, ceramics, etc.)	Exhibited or published at my college one or more works of art, such as drawings, paintings, sculpture, etc. Had drawings, photographs, or other art work published in a public newspaper or magazine Sold one or more works of art, such as drawings, paintings, sculptures, ceramics, etc.
Science	Participated in a National Science Foundation summer Program for high school students Won a prize or award of any kind for scientific work or study Gave an original paper at a scientific meeting sponsored by a professional society	Took part in the Undergraduate Research Participation program (URP) of the National Science Foundation Received a prize or award for a scientific paper or project Gave an original paper at a convention or meeting sponsored by a scientific society or association

TABLE 2. (*Continued*)

Area	*High school items*	*College items*
Music	Composed music which has been given at least one public performance Performed with a professional orchestra Received a rating of "good" or "excellent" in a state music contest	Composed or arranged music which was publicly performed Have been paid for performing as a professional music teacher on a continuing basis Attained a first division rating in a state or regional solo music contest
Writing	Had poems, stories, essays, or articles published in a school publication Had poems, stories, or articles published in a public newspaper or magazine (not school paper) or in a state or national high school anthology	Had poems, stories, essays, or articles published in a college publication Had poems, stories, essays, or articles published in a public (not college) newspaper, anthology, etc.
Speech and Drama	Had leads in high school- or church-sponsored plays Had minor roles in plays (not high-school- or church-sponsored) Placed first, second, or third in a regional or state speech or debate contest	Had one or more leads in plays produced by my college or university Had one or more leads or minor roles in plays not produced by my university Placed second, third, or fourth in a contest in speech, debate, extemporaneous speaking, etc.

included. In all, they used 130 predictors. In the results, high school grades and the SAT did not appear among the predictors selected by a stepwise multiple regression program as predictors of college accomplishment. One interesting feature of this study was that there seemed to be little point in distinguishing between all achievements and rare achievements:

> The most notable finding . . . is that achievement in high school or daily activities, interests, and involvements which are related to achievement (Potential scales) are the best predictors of achievement in college. Expressed goals, such as grades a student expects to receive in college or 'making a contribution to scientific knowledge,' are next in predictive efficiency. These two trends are followed by a variety of measures of lesser usefulness—the Indecision Scale, intellectual resources in the home, number of competencies, etc. Of special interest, the Scholastic Aptitude Test (Verbal and Mathematical scales) failed to enter the multiple correlations at this high level of aptitude. (p. 64)

Nichols (1966) subsequently followed a similar strategy. He developed new scales from items taken from the Adjective Check List, the Vocational Preference Inventory, the California Psychological Inventory, and an objective behavior inventory, consisting of a listing of 326 hobbies, sports, leisure time activities, interactions with other people, etc.

A critical study by Holland and Richards (1965) is important, not only because it shifts the research activity to the American College Testing Program, but because it is based on a large, diverse, and typical sample of college freshmen. The sample of 3,770 men and 3,492 women included students from a wide variety of ability levels and appeared to be a reasonably representative sample of the national college freshman population. The colleges included a wide variety of institutions. A new type of measure was used, which assessed students' competencies in a wide variety of areas. Students checked from a list of 143 those activities which "you can do well or competently." The assumption underlying these scales is that a large number of competencies is conducive to achievement generally and that competencies in a particular field are conducive to achievement in the same field. Typical items for this list included: "I have a working knowledge of *Roberts' Rules of Order,*" "I can make jewelry," "I can read blueprints." The number of activities checked equals a student's range, or total number, of competencies. Reasonably reliable scales were also developed in eight areas of competency, such as scientific, leadership, art, etc. The reliabilities (K–R 20) of the achievement scales were considerably higher than those used in the NMSC studies, ranging from .72 to .84 for men and .65 to .81 for women.

In general, the correlations between academic measures and the achievement scales were significant but low, averaging .04. The highest correlation between ACT test scores and achievement was .18. The highest between

grades and achievement was .21. In contrast, the competency scales correlated with many of the achievement scales at a moderate level.

The researchers also examined the possibility that the basic relation between academic ability or performance and the socially relevant accomplishments examined in these studies is curvilinear. That is, the correlations are low because only the very able are truly able to achieve; i.e., the distribution would be so skewed to the high-ability end that the correlation would appear low. Holland and Richards compared eta coefficients with the Pearson product-moment correlations, examined the scatter plots, and found no evidence for the idea.

Finally, they computed biserial correlations between the items in the achievement scales and the four ACT scores and average high school grades. "This analysis was important to perform for several reasons: since the scales of nonacademic accomplishment contain many low level accomplishments, they may assess quantity rather than quality of accomplishment." The median correlations between ACT scores and achievement was .03 for men, and .05 for women; the range was from $-.15$ to .22 for both sexes. The median correlation between grades and achievements was .03 for men and .05 for women; the range was from $-.13$ to .36 for men and $-.11$ to .32 for women. Some 90% of the correlations fell between $\pm .15$.

In 1966, Richards, Holland, and Lutz attempted to develop revised scales of college level accomplishments in the six areas assessed in the earlier studies, and to develop new scales in the areas of social science achievement, humanistic-cultural achievement, business achievement, social participation, social service achievement, and religious service. They also developed a scale of recognition for academic accomplishment. The scales were administered to freshmen in 6 colleges, sophomores in 31 colleges, and seniors in 12 colleges. Expectedly, means on the 10-item scales increased from class to class. The median reliability coefficient among men was .65 for freshmen, .66 for sophomores, and .71 for seniors. The corresponding figures among women were .62, .59, and .70.

The researchers found that there were low relationships between college grades and the accomplishments assessed in both the six areas studied in earlier studies and the new areas in all three samples. In contrast, grades were correlated with the five-item scale, recognition for academic accomplishment; the correlations ranged from .30 to .46. This result is important because it suggests that neither the brevity nor the skewness of the other accomplishment scales produced the lack of relationship with grades.

In a subsequent study, Richards, Holland, and Lutz (1967) again studied the freshman and sophomore samples, this time examining the predictors of college level accomplishments among information collected from the students when they were applying to college. Neither high school grades nor

ACT test scores had correlations as high as .20 with any college accomplishment in any nonacademic area in either sample.

To examine the possibility that this lack of correlation was due to a procedure which had grouped together students from many different colleges, the researchers also computed the correlations of ACT test scores and high school grades with college achievements of males at each individual college in the sophomore sample. The median correlations were very similar to the correlations for the total sample. The typical correlation was close to zero, although the correlations between ACT tests and nonacademic college accomplishments ranged from −.53 to .41 and the correlations between high school grades and nonacademic college accomplishments ranged from −.49 to .31. There was no systematic relationship between the size of the correlations and characteristics of the colleges. For example, there was no trend for the correlations to be positive in selective colleges and negative in unselective colleges.

A meta-analysis (following procedures suggested by Glass, 1978) of the results of the NMSC and the American College Testing Program results that were based on "typical" samples of college students was performed. It examined between 34 and 60 correlations of academic ability tests and grades with the accomplishment scales that were reported in these studies. As shown in Table 3, the results showed median correlations between leadership and grades of .15, and leadership and test scores of .08; between science and grades .07, and between science and test scores .09; between writing and grades .08, and between writing and test scores .14; between dramatic arts and grades .04, and between dramatic arts and test scores .04; between music and grades .00, and between music and test scores .06; and between art and grades −.03; and between art and test scores .03. Thus, in general, there are low positive relationships between academic ability, grades, and extracurricular accomplishment in leadership, science, and writing, but not in the other areas.

Why would this be so? Students engage in activities for a variety of reasons, related to their needs, their personalities, and their interests. The degree of their participation can be influenced by major fields, classes, professors, friends, and residences, to name only some obvious influences. These personal characteristics and situational variables work independently of academic ability and may well be more influential. (In fact, some NMSC and ACT studies suggest that both personal and institutional characteristics do have consistent influences on accomplishment.) For example, students aspiring to be television writers may write stories or plays no matter what their academic ability. Or a student with a strong need for self-expression may audition for and obtain parts in plays. A student who has worked in chemistry from an early age may conduct an experiment. A music major

TABLE 3. Summary of Meta-Analysis of the Relationship between GPA, Academic Ability, and Nonacademic Accomplishment

	Median Correlation	*High Value*	*Low Value*	*Modal Categories*	*N of Correlations*
A. With GPA					
Leadership	.15	.27	-.04	59% between .11 and .25	59
Music	.00	.16	-.06	79% between -.05 and .04	58
Drama and speech	.04	.12	-.06	75% between .01 and .10	57
Art	-.03	.13	-.10	83% between .00 and -.10	60
Writing	.08	.17	-.05	82% between .01 and .15	60
Science	.07	.19	-.08	70% between .01 and .10	57
B. With Academic Ability					
Leadership	.08	.20	-.01	59% between .01 and .10	34
Music	.06	.13	-.07	71% between .01 and .10	34
Drama and speech	.04	.18	-.05	62% between .01 and .10	34
Art	.03	.21	-.06	62% between .01 and .10	34
Writing	.14	.22	.04	71% between .11 and .20	34
Science	.09	.20	.00	68% between .01 and .15	34

may have to compose and perform a composition as part of a class requirement. A student in a speech class may be encouraged to enter a debate contest for extra credit. A professor may encourage a student to submit a poem for publication, and another professor may encourage a student to cooperate in the writing of a paper. A student's friends may encourage her to run for class office. A residence may emphasize participation in campus clubs or political activity. None of these personal or situational spurs to participation and accomplishment necessarily has anything to do with academic ability.

Thus, it is not that academic ability is irrelevant in accomplishment, but rather that it is one among many factors influencing college attainment. One of the most important of these is simple participation—students who do not enter contests cannot win them—a variable probably most influenced by interests and needs. perhaps the next most important is the degree of participation—students who have roles in many plays are more likely eventually to play a lead than students who have roles in only one or two. The degree of participation is probably most influenced by persistence, enjoyment of the activities, and encouragement received.

To summarize, the correlational studies of grades and academic ability tests in relation to scales of accomplishment show a small relationship. The small relationship does not seem due to unreliability, skewness, or other statistical defects of the scales. The samples of students and adults represented a wide range of ability in a variety of types of schools, colleges, and situations and included individuals ranging from high school students to college alumni. The patterns of correlations of accomplishments and academic potential with personality, interest, value, self-concept, and activity variables also suggest that there is a small relationship.

Although the results are very consistent, some individuals may still question them on the grounds that the correlations obscure distinct differences between the highly academically able and the average person. They argue that one needs to examine the distributions of accomplishment at different levels of academic ability.

There have been several studies of the frequency of accomplishment at several levels of academic ability or grades. They have been of two types; comparison of groups and simulated selection studies. The first of these studies was conducted by Astin (1964), who compared 334 Merit Scholars with an unselected sample of entering college freshmen at 248 colleges who were matched with the scholars on socioeconomic background. The comparisons of high school accomplishments revealed considerable superiority of the Merit Scholars in science and writing, slight superiority in leadership, and approximate equality in drama, art, and music. These results appear to be in disagreement with the correlational results. However, it should be

recalled that half of the Merit Scholars were selected by a committee which not only examined their test scores but also studied their high school accomplishments. In addition, a number of scholarships were awarded because the students showed promise of exceptional achievement in a particular area or because they were judged creative. In addition, a large number of the remainder were awarded according to criteria stipulated by a sponsor, which sometimes included exceptional accomplishment.

Perhaps a less biased comparison was made by Baird (1968), who compared the college accomplishments of a typical cross section of students (described elsewhere by Baird, 1969a), with the accomplishments of the very bright National Merit Finalists described by Nichols and Holland (1963). Both groups reported their accomplishments at the end of their freshman year. Comparisons were based on the percentages reporting 35 specific accomplishments. Baird found that, in general, there was very little difference between the two groups.

In a second substudy, Baird compared the number of high school achievements of bright and average students, using data from the Michigan Scholarship Program, which regularly tests a large number of Michigan high school students and uses the ACT test battery as a basis for considering students for scholarships. Only students with an ACT composite score of 22 or above were considered eligible for scholarships. Baird compared students who were eligible with those who were not. The mean ACT composite score of 14,424 eligible students was 25.5, approximately the 86th percentile of students enrolled at ACT-participating colleges (American College Testing Program, 1973). The mean of 10,680 students who were not considered eligible was 18.2, approximately the 35th percentile on national norms. Baird compared the number of high school achievements for the two groups using simple analysis of variance. The researcher also calculated Hays's (1963) omega squared (ω^2), a statistic (similar to the intraclass correlation coefficient) which assesses the strength of an association between variables by estimating the proportion of variance in a dependent variable accounted for by the independent variable. The distributions of the number of accomplishments were very similar in both groups, although there were small significant differences favoring the high-ability group in writing, leadership, and science. Although these differences were significant, the omega-squared values indicated that in no case did academic ability account for as much as 1% of the variance in accomplishment.

Elton and Shevel (1969) followed yet another strategy to study the relationship between academic ability and nonacademic accomplishment, comparing the high school accomplishments of students who were one standard deviation above the mean on both the ACT English and the mathematical

scores, with the accomplishments of those who were below average on both and of those who were average on both. The high-high group composed about 3% of the sample, as did the low-low group, and the average group about 60%. The sample was a 3% random sample of students who completed the ACT examination in 1966–1967. When the comparisons of the individual accomplishments of the men and women in the high-high group with the low-low group for men and women are combined, the results showed differences favoring the high-high group on 20 accomplishments, differences favoring the low group on 14, and no differences on 62. When the high-high group was compared with the average-average group, the comparisons favored the high group on 34, the average group on 9, and showed no difference on 53. When the average-average group was compared with the low-low group, the comparisons favored the average group on 8, the low group on 8, and showed no difference on 80. Thus, there seemed to be slight evidence for a relationship between academic ability and accomplishment.

Subsequently, Werts (1967) calculated the proportion of students at several grade levels who had demonstrated accomplishment in 18 different areas of attainment. Wert's sample was 127,125 students who had completed a survey of their plans and high school activities when they began their first year of college. Students with high high-school grades tended to have somewhat more accomplishments than the students with low high-school grades. For example, among males, 14.4% of the C students versus 31.8% of the A students had had a lead in a school play.

Holland and Richards (1967a) replied to Werts by reanalyzing Wert's data to show "what you miss by various selection rules as well as what you get. . . . By reanalyzing Werts' data . . . we created a single table that shows what percentages of students with various kinds of achievement are eliminated by the use of various grade levels as selection scores" (pp. 205–206). They found that

> The selection of only A + or A students (a selection rule that will admit nearly all students in the top decile of grades) will result in the elimination of 74–93% of all students with various kinds of nonacademic accomplishments. To take another more concrete example, if you only select the A or A + students (about the top decile of academic talent), you would get 1,843 class presidents, but you would miss 11,096 class presidents. . . . In short, the use of grades as an efficient sign for the selection of multitalented persons is not warranted by the Werts data. (p. 206)

Holland and Richards then went on to reply to Werts and other critics by presenting evidence to make four points which oppose contentions by their critics: (1) the small percentages of students with nonacademic accomplishments do not present a misleading picture of the actual relationships between academic and nonacademic accomplishments; (2) the lack of relation-

ship is not due to a narrow range of talent; (3) the lack of relation is not due to statistical artifacts; and (4) one cannot use academic criteria for selection and hope to select a group of students who will achieve in nonacademic creative areas. On the last point, Holland and Richards open the question of the comparative consequences of using academic ability and nonacademic accomplishments for selection purposes. Earlier, Nichols and Holland (1964) had studied these consequences in a sample of National Merit Finalists. Information on earlier performance had been collected, and the criteria were academic and nonacademic achievement in college. They examined nine alternative methods for selecting students, including selecting on ability tests, on grades, and on accomplishments. Their conclusions included:

> (a) Additional selection on aptitude using either the same or a different test does not appreciably improve selection for high-level college performance. (b) Selection on the basis of high school rank produces students who demonstrate superior academic performance, but not necessarily other kinds of achievement. (c) Selection on the basis of a broad range of high school achievements results in a broad range of achievement in college without lowering the level of academic performance. (p. 33)

Subsequently, Wing and Wallach (1971), at Duke University, used some of the Holland and Richards (1965) scales to examine the types of classes one would obtain if one selected on the basis of SAT scores alone, on the basis of high school rank alone, on the basis of both SAT scores and high school rank, and on the basis of creative accomplishments. Criteria were the students' high school accomplishments and personal characteristics. Expectedly, the SAT strategy selected students with higher high school rank, the high school rank strategy selected students with high SAT scores, and the use of both selected students high on both. None of these strategies was particularly successful in obtaining a class with many high school accomplishments. Using high school creative accomplishments as the admissions criterion, Wing and Wallach found that a class high in such accomplishments would be slightly higher than the total population on SAT scores and high school rank. Although Wing and Wallach provided extensive comparisons of the characteristics of students about whom the admissions decisions of the strategies disagreed, they did not show the characteristics of the students who would be rejected by each strategy by itself.

The Wing and Wallach study was criticized in various quarters for the restriction of range of academic talent (average SAT verbal and mathematical scores of the applicants were close to 600, the accepted students close to 650), for the fact it was confined to a single institution, and for the lack of follow-up data, that is, the performance of the admitted students in college.

Earlier than the Wing and Wallach study at Duke, Baird and Richards

(1968) examined the effects of various selection strategies in a large sample of students in 35 diverse colleges. In contrast to the Wing and Wallach study, analyses showed their sample to include a wide range of academic ability and to be representative of students at these colleges. Furthermore, the success of the selection strategies was evaluated by college criteria of the number of college accomplishments in six areas, college grades below C, college grades of A, and the percentage who had dropped out. The strategies were: (1) admission only on the basis of grades; (2) admission on the basis of high school creative accomplishments; and (3) admissions on both. First, the characteristics of entering classes which would be admitted by the strategies were examined. These analyses showed that the use of grades to select students would result in a class of students who would make passing grades (but few who would make A grades in college), who would not drop out, and who would not be more (or less) likely to achieve in nonacademic areas. Admission on creative accomplishments would result in college classes that would include many students who would write stories and essays, develop their own science experiments, create their own music, take part in college and noncollege plays, submit works of art to art contests, and run for campus offices. The students selected by this strategy would be also somewhat less likely to drop out but were not more (or less) likely to have good grades.

When the selection strategies were compared on the basis of the number of college achievers who would be eliminated by the strategy, Baird and Richards found the following results: stringent selection on the basis of high school grades would result in the elimination of most students who would have college accomplishments in leadership, art, music, speech and drama, writing, and science. In addition, the selection of students with high school grades of B+ or above would also result in the elimination of three-quarters of the students who would obtain passing grades in college. The results for the strategy of using high school creative accomplishments for selection are complicated because the consequences vary from area to area. However, it is clear that stringent selection on nonacademic accomplishments also eliminates many nonacademic as well as academic achievers in college.

Other Studies of Students

We have concentrated on the NMSC and the ACT studies to this point because they form a continuous body of inquiry. However, there are several other studies involving students that have examined the relationship between academic ability and accomplishment and that used other measures and often concentrated on other questions.

For example, Milgram and Milgram (1976) used a creativity test and a

variant of the accomplishment criteria just described. They studied an entire high school class (60 boys and 85 girls) in Tel-Aviv. The criteria were self-reports of accomplishments adapted from Holland and Richards in nine areas ranging from science to sports. All were lumped together in three different scoring systems. High scorers scored high on an adapted version of the Wallach and Kogan (1965) creativity battery, but not on IQ or school grades. They also found that quantity and quality of accomplishment could not be separated.

Similarly, Cropley (1972) administered six creativity tests plus an IQ test to seventh graders; five years later, 111 of these students were questioned concerning their art, drama, literature, and music attainments, using the Holland and Richards scales. No significant correlations were found between IQ and the criteria among girls, and only one (.32 with literature) among boys. Cropley found some low positive correlations between attainments and creativity tests.

Skager, Shultz, and Klein (1965) developed an instrument similar to the Holland and Richards scales, the Independent Activities Questionnaire. They derived scores on the number of high school accomplishments (quantity) and then judged the quality of the accomplishments of 142 male state university entering freshmen and 150 male technological institute freshmen. In neither sample was the SAT-verbal score, SAT-mathematical score, or high school rank related to quality or quantity. However, when the samples were combined, there were small correlations between quality scores and the SAT, chiefly because the technological institute students had higher SAT and quality scores.

In a study by Locke (1963), 122 high school juniors and seniors attending a Cornell University summer NSF program were given a large battery of tests which were factor-analyzed to yield 11 factor scores. Criteria were classroom achievement, as represented by grades and teacher ratings, and out-of-class achievement, as represented by ratings of the amount and quality of independent scientific work done and teachers' comments about the students. Classroom achievement was predicted by measures of vocabulary, self-control, and high socioeconomic status versus independence. Out-of-class achievement was predicted by school and city size, creative energy, independence, and originality. Vocabulary and general reasoning measures were unrelated to out-of-class achievement in the total sample.

Using somewhat different criteria, James et al. (1972) obtained faculty ratings of 813 high school students for the areas of dance, music, theater, and visual art. They also used checklists of creative activity and recorded the number of awards in arts obtained by each student. Although no detailed results are reported, they concluded, "The correlations between the art criteria and academic GPA varied from nonsignificant to moderate and signifi-

cant. In general, it appeared that somewhat different subsets of abilities were required for successful performance in the two fields."

The five studies just reviewed were general research studies, concerned chiefly with studying the relationships among variables. The next three were related to the practical concerns of scholarship programs.

Datta (1967) studied high school senior applicants to the Westinghouse Science Talent Search (STS) who had submitted a research project that was judged for "creativity and potential creativity" by the refined methods of the search. Only students who scored above the 80th percentile on a scientific aptitude test were included. Five hundred thirty-six of these students were divided into three groups, differing on the rated creativity of their project. There were no significant differences between the groups on the SAT verbal or SAT mathematical scores. It should be noted, however, that the mean SAT scores were quite high.

In a later study, Parloff et al. (1968) compared 266 participants in the STS whose reports of an independent research project were judged creative with 672 whose projects were judged less so. (All scored at the 80th percentile or higher on a science aptitude test.) There were no differences on SAT verbal or mathematical scores, high school grade average, social class, or birth order. There were, however, differences on some personality scales.

Edgerton (n.d.) examined data from the 1968–69 STS for the Westinghouse Scholarships and Awards program. Students around the country submitted an independent research project, a report of more than 1,000 words, a personal data blank, and a high school transcript. They also completed a science aptitude examination. From the 2,356 seniors found eligible, an honors group of 300 was selected, and 40 scholarship winners were chosen from the latter group. Four selection models were compared: (1) a model placing primary emphasis on quality of research project ("a basis of actual performance analagous to that of adult scientists"); (2) a model using the successive hurdles of academic achievement in high school, scores on the science aptitude examination, and then the project; (3) a model using examination scores and academic achievement only; (4) a model using a composite of attainment in all areas. Results were "that two-thirds of the students chosen for their Scientific Performance would not have high enough scores if Academic Achievement had been the sole criterion. And two-thirds of those chosen on a basis of Academic Achievement had such low ratings of their Project Reports and Personal Data Blanks that they were not included in the Scientific Performance. . . . Since the overlap among the students chosen by these two means was relatively small, it strongly suggests that evidence of scientific talent as indicated by actual scientific performance is only partially related to academic achievement."

Similar results were obtained by Schmidt (1973) who used six measures to

predict the standing on seven criterion measures of creativity among 105 first-year architecture students. A measure of academic success was not related to the criteria.

Getzels and Czikszentimihalyi (1975) administered a six-hour battery of tests to 179 students at the School of the Art Institute of Chicago, one of the leading art schools in the country. In addition, several hundred additional Institute students completed part of the battery of tests. As a group, the art students scored close to the average for college students on the cognitive tests used, but differed markedly from college averages on measures of values and personality. Mixed results were obtained when test scores were related to grades in studio art courses and to teachers' ratings of the students' originality. For example, among female students in applied art, there were some positive relationships between art grades and perceptual and cognitive tests, but among the male students, these relationships did not hold and were sometimes negative. The authors mention a positive correlation of .52 between spatial visualization and art grades for female students; for the male students it was $-.32$. In any case the authors concluded that traditional academic ability is of relatively little importance in art.

Similar conclusions were drawn by Burkhart (1967), who reviewed various studies of the relationship between artistic performance in school and academic success and measures of intelligence.

Finally, Mednick (1963) studied 43 University of Michigan and Northwestern University graduate students in psychology. They were rated on Taylor's (1963) research creativity scale by their research advisors. Neither GPA nor Miller Analogies Test scores were related to the ratings, although a measure of creative thinking was.

Conclusions

The studies just described present evidence that is in general agreement with the evidence of the NMSC and ACT studies. Overall, whatever the purpose of the study, and however the sample was selected, the results have generally been the same: very small relationships between academic ability and accomplishment in the high school or college years.

The results of these various studies do not necessarily imply the elimination of academic tests and grades as admissions criteria. These are by far the most efficient predictors of academic performance in college, and academic performance is the most important part of students' collegiate careers. The point is simply that colleges interested in other kinds of performance should look for evidence of potential for those types of performance among their applicants. Since earlier accomplishment in an area is by far the best predictor of subsequent accomplishment, attention should be devoted

to these attainments. And this attention is certainly consistent with the current admissions situation. That is, more and more students delay entry to college after high school, leave college for a few years, and seek unusual work or other experiences during their breaks from their studies. In addition, more students work during school and college, and more older students are entering college. Many of these students have had educationally valuable experiences outside the classroom, for which they received no credit. Many schools and colleges have also begun a wide variety of off-campus programs of independent study, work experience, public service, and so on. In this way, many students have opportunities they would not otherwise have had to develop and demonstrate their talents.

Clearly, tests and traditional undergraduate transcripts do not provide adequate means of recognition for these kinds of learning and accomplishment. For these reasons, it is important to find ways to assess the accomplishments of students.

Another purpose in assessing students' accomplishments is to select students who are likely to be productive, to be creative, to provide leadership, and to make a contribution to their fields. Many admissions committees, faced with large numbers of applicants and dwindling funds, feel the need for some way to assess the high-level, noninstitution-sponsored accomplishments of students. They wish to have some way of selecting students who will be outstanding students and who will eventually contribute most to society. As the review of research indicates, the most efficient information for predicting future accomplishments is data on previous accomplishments. The studies reviewed show that the best predictors of future high-level, real-life accomplishment in writing, science, art, music, and leadership are similar accomplishments, albeit at a lower level, in previous years. In fact, as with all other behavioral and scientific prediction, which is based on the consistency of the same or similar phenomena over time, the studies indicate that the most effective predictor of high-level accomplishment is past high-level behaviors of the same or similar types. People who have been outstanding in a wide variety of areas in science, literature, creative arts, and public affairs have been shown to have had accomplishments in those areas in their high school and college years. The institution that wishes to have graduates who will be outstanding in their fields in the future might well consider the previous accomplishments of their applicants.

To date, information about past accomplishments has proved to be a far better predictor of high-level accomplishment than measures of ability, interests, or personality. As the present and an earlier review (Baird, 1976a) both indicate, scales of real-life accomplishments can be constructed that are reliable, usable, and seldom faked. They can be used in selection decisions in a variety of ways. They seem particularly useful when there is a need to

assess talents somewhat removed from academic ability, such as artistic capacity, musical skill, ability to write expressively and forcefully, dramatic power, and the intuition needed to devise a scientific experiment. As these examples suggest, the assessment of talent is more difficult in some areas than in others, and, consequently, the predictive power of the variables will vary from area to area. In any case, these measures cannot replace measures of academic talent; they simply provide indications of capacity, in and out of class, in other areas that are useful for specific purposes. The range of talents that institutions consider in their applicants could be greatly expanded if they used these measures. Thus, an institution could not only select students who will get good grades, but students who will be good organizers of research, leaders in political and nonpolitical organizations, good writers, and inventive experimenters.

Another important reason for developing measures of in- and out-of-class activity is that the student applying for study has a right to be able to present his or her skills, talents, and achievements to selection committees. As recommended by the College Board's Commission on Tests, students should have some choices in the picture of themselves that selection officers see. And there is a further positive outcome of the inclusion of this sort of information. The students who complete a form that asks for their personal accomplishments may feel that they are being taken more seriously and that they have had a chance to present their best side.

GENERAL SUCCESS AND SOCIOLOGICAL STUDIES OF CAREER ATTAINMENT

The studies reviewed to this point have used relatively specific criteria of accomplishment within fields or specialized areas of activity. It is possible that the generally low relationship between these criteria and academic ability may be attributed to their specificity. That is, it may be that attainment in specific roles or positions is so narrow that the full force of academic ability cannot be seen. However, it is possible that if one were to look at success or accomplishment across positions or occupations, one would find that academic ability plays a large role. Perhaps more global criteria such as general success, occupational status, and personal income across a wide range of ability would be more appropriate. Although these criteria are obviously more ambiguous and problematical, they are the only ones that can apply across a heterogenous group of careers.

There have been three general categories of studies: long-range follow-ups of college students, studies of occupational attainment, and sociological studies of career attainment.

Long-Range Follow-Ups of College Students

Researchers at several institutions have followed up samples of their alumni to determine their level of "success." The criteria have differed, and the samples have been divided in various ways. For example, Jones (1956) examined the careers of graduates of the University of Buffalo from the classes of 1929 to 1941 (the average was 20 years after graduation) in relationship to the data available for them as college freshmen. Self-reported income and professors' ratings of their success were the criteria. For income among the arts and science majors, college grades correlated .34 (significant) with income, test scores .04 (nonsignificant); among business majors, grades correlated −.04 (nonsignificant) and tests −.29 (significant). Professors' ratings of arts and science majors' success were correlated with grades .30 (significant) and tests .24 (significant); for business majors, grades were correlated .20 (significant) and tests .18 (nonsignificant). When income was used as the criterion within the occupational groups of (1) science (including M.D.s), (2) social service (including law), (3) education, and (4) business, college grades were not significantly related in any group. However, tests were positively related in education (.48), and negatively related in business fields (−.23). When only lawyers were studied, neither grades nor tests were related to income. Jones's study suggests some of the problems of using general levels of "success" as criteria. Obviously, some occupations have higher average incomes than others, and entrance to some, such as law and medicine, are dependent on high grades and test scores.

Another complexity is suggested by a study by Elder (1968). For 63 men, Elder found that IQ scores obtained in 1938 predicted occupational status ($r = .42$) and educational level ($r = .50$) in 1958 for middle class men, but they did not predict either of these for working class men. In contrast, the Strong Vocational Interest Blank "occupational level" score did predict in the latter group, but not in the former. Elder's study illustrates the importance of an individual's social class. Social class influences high school graduation, entrance to college, attrition during college, and entrance to graduate and professional school (Baird, 1976b).

For 619 male university graduates, Lewis (1975) found a small, but significant, association between occupational success and the Iowa Placement Tests. The University of Iowa administered the tests to students admitted in the academic years 1948–49, 1954–55, and 1959–60. Lewis followed them up in the late 1960s. Occupational success was defined by Roe's (1966) system into three groups. The distribution of occupational success for the half lower in ability was level 1 (highest), 9%; level 2, 72%; level 3, 19%. The corresponding figures for the highest quartile were 24, 66, and 10. For those between the 50th and 74th percentile, the figures were 16, 74, and 10.

Wolfson (1976) followed up after 25 years 306 women who had attended the University of Minnesota during 1933–36. They were placed into five career categories of career accomplishment ranging from "never worked" to "unusually high accomplishment"; 29 variables were studied. The variables that discriminated most clearly among the vocational patterns were those related to marriage and education. Graduation from college, a vocational major, attendance in graduate school, and unmarried status were most characteristic of women with the highest vocational patterns. The Minnesota Scholastic Aptitude Test did not discriminate among the groups; the Minnesota College Aptitude Test did, but the groups were not arranged in any meaningful order. The most successful career group had the lowest scores. However, it should be noted that women during the 1930s, 40s, and 50s probably had to face a good deal of sex discrimination, which probably affected the results in many ways.

Nicholson (1970) examined the later success of members of the Brown University classes of 1950, 1951, and 1952. "Success" was defined as meeting a number of criteria: reputation for academic or research accomplishment, contribution to the national community, income, etc., as judged by a panel of the alumni of each class. Similar judgments were made of lists of Brown alumni from the same classes who were included in *Who's Who* and other national biographies, or whose biographies in the *Brown Alumni Monthly* seemed to meet the criteria. Altogether, the 1,105 verbal high scorers (SAT verbal scores above approximately 490) were rated successful in 26% of the cases; the 1,022 verbal low scorers (below 490) were rated successful in 23% of the cases. Later analyses comparing the mean scores of the successful and unsuccessful groups showed no significant differences on the SAT verbal and SAT mathematical scores for either the alumni who were veterans or those who were nonveterans. However, among nonveterans, high school class rank and high school average were higher for the successful alumni. First semester college GPA was also higher for the successful students in both groups.

These studies obtained inconsistent results, suggesting a small relationship between academic ability and success. However, this possibility has to be weighed against the effects of social class, years of education, type of occupation, and degree obtained. Is there any way to sort out these influences?

Studies of Occupational Attainment

Probably the most frequently cited evidence about the relationship between academic ability and occupational attainment is Harrell and Harrell's (1945) study of the ability of World War II enlisted men who had been in different preinduction occupations and Thorndike and Hagen's (1959) study

of the later occupations of 10,000 World War II Air Force cadets. The Harrell and Harrell study reported the mean, median, standard deviation, and range of Army General Classification Test scores by male inductees' civilian occupations. For example, they showed that accountants averaged 128.1, with a range of scores from 94 to 157; mechanics averaged 106.3, with a range from 60 to 155; and teamsters averaged 87.7, with a range from 46 to 145. Although these results suggest a substantial general relationship between test scores and occupational attainment, they are limited by the fact that they are retrospective; i.e., the scores of various occupations may have been the results of educational or other experiences rather than differences in innate ability.

The Thorndike and Hagen results deal with some of these problems by obtaining the test data prior to data about subsequent occupations (12 years later). In addition, Thorndike and Hagen used five scores—general intellectual, numerical fluency, visual perception, mechanical, and psychomotor. These scores allowed them to construct profiles for the individuals in each occupational group. For example, the cadets who eventually became lawyers had above-average scores on general intellectual capacity and numerical capacity, but had below-average scores on the mechanical and psychomotor tests.

In general, the scores of cadets in different groups were about as expected. For example, the cadets who later were college professors, engineers, physicians, and scientists had scored high on the general intellectual score composites. Those who later became managers, pharmacists, treasurers, accountants, and securities salesmen scored high on the numerical composite. Those who later became architects, artists, surveyors, and radio-TV repairmen scored high on the visual perception composite. Those who later became airline pilots, carpenters, electricians, and wood-carvers scored high on the mechanical composite. Cadets who later became appliance mechanics, machinists, firemen, and plasterers scored high on the psychomotor composite. Of these measures, the general intellectual composite seems closest to a measure of general academic ability, and the distribution of later occupations on the measure is close to what one would expect. For example, the highest-scoring groups included cadets who later became engineers, physical scientists, college professors, social scientists, physicians, treasurers, office machine mechanics, and architects. The groups that were very close to average included cadets who later became buyers, artists, clerks, draftsmen, laboratory technicians, credit managers, and real estate salesmen. The lowest-scoring groups included cadets who later became production-line assemblers, earth movers, crane operators, welders, linesmen, painters, pumpmen, and bus and truck drivers. In general, these groupings mirror the usual status and income rankings of the same occupations.

Thorndike and Hagen also sought to predict success within occupations,

defined as reported income, vertical progress within the occupation, stability in the occupation, work satisfaction, personal sense of success, number of individuals supervised, and length of time spent in the occupation. In this case, however, the number of significant correlations between test scores and the criteria was close to the number expected by chance. For example, of the 385 correlations between the composite scores and income, for 77 occupational groups, only 24 or 6.2%, were significant at the .05 level. Thorndike and Hagen concluded that the null hypothesis seemed adequate to account for their results (p. 45).

They noted, however, that several factors worked against their finding significant validity coefficients within occupations. Especially important was the fortuitous nature of the employment situation. First, they pointed out that the criteria were imperfect, describing many of the same difficulties noted in earlier pages in this review. Second, their predictors included only tests of ability and a brief biographical form. Measures of personality, interests, social skills, and the like were not included. Third, their sample was preselected, representing roughly the top half of a high school graduating class. Finally, the heterogeneity of work within occupations makes prediction difficult. A lawyer or an accountant may have a small practice in a small town or may be employed as a senior official in a Wall Street firm—situations that are probably more due to personal preferences and personality than to differences in ability.

In any case, Thorndike and Hagen present strong evidence that ability test scores are related to the occupational outcomes of individuals in expected ways.

As impressive as these studies are, they have been questioned because they do not control for the social class background of the subjects, the influence of education, and other personal characteristics that may affect the results. That is, do the test scores reflect basic abilities that help to cause the differences in occupational attainment, or are the test scores the result of favored social position, educational opportunities, etc.?

Analytical Sociological Studies

In an attempt to answer this last question, a number of sociologists have analyzed the attainment process. They have attempted to estimate the influence of social class, ability, education, family influence, and other variables on the occupational status or income attained by people in the United States.

The sociological studies that have attempted to analyze the achievement process have used a variety of samples and measures. The criteria have generally been occupational status (especially as defined by Blau and Duncan, 1967) and income. Occupational status means the esteem in which the

occupation is held. For example, a physician would obtain a high score, a garbage collector a low score. In other analyses, occupations are grouped into categories such as professional, high managerial, etc., and the category is assigned a score. Income has been treated as a linear variable or as a transformed variable; it has sometimes referred to salary alone and at other times to total income. However, whatever the details of the definitions, the basic constructs of occupational status and income are the same and reasonably clear.

The various studies have examined different career periods and have obtained different data at different times. However, whatever the differences, the basic technique that has been used is path analysis. This method attempts to create models of the influence of one variable on another and produces estimates of the amount of that influence that may be due to another intervening variable. This is much more informative than simple zero-order correlations.

A great deal of this research shows that people who score high on tests of academic ability tend to obtain higher status jobs and earn more money than people who score lower. However, people with different social viewpoints interpret this finding differently. Liberal views emphasize the role of social class in determining scores and the presumed "biases" in the tests themselves (e.g., Block and Dworkin, 1976; Bowles and Gintis, 1976; Kamin, 1974). More conservative views emphasize the importance of academic ability per se and the necessity of so-called "middle class skills values" for a technological society. Crouse (1979) has reviewed the empirical results bearing on the effects of academic ability, by reanalyzing the data from Project Talent, Sewell and Hauser's (1975) sample, the Equality of Educational Opportunity sample (Alexander, Eckland, and Griffin 1975), the Kalamazoo sample (Olneck, 1976), the SRC sample, and the Armed Forces Qualifying test sample (Jencks and Rainwater, 1977). Crouse attempted to control for such variables as parental social class and to at least estimate the role of ability in influencing the extent to which adolescents are enrolled in college preparatory curriculums, earn high grades, receive parental and peer encouragement of their college plans, and receive the attention of their teachers. Crouse concludes that even after controlling for social class, curricular placement, etc., more than half of the observed correlation between test scores and educational attainment remains. He leaves open the question as to whether this is due to merit or to causes that are "unfair" to adolescents with low ability. Examining the relationship between ability and earned occupational status, Crouse finds that there is an important effect of ability, but that "60 to 80 percent of the effect is explained by the amount of schooling the individual attains. . . . Men who fail to convert their ability advantage into additional schooling do not have much of an occupational advantage over men with lower scores."

The effects of ability on earnings were as follows. (1) Even controlling for family background, a 15-point test-score difference is associated with a 17% difference in annual earnings in a sample of brothers. (2) The effects of test performance on earnings increase with age. (3) Although differences in earnings are partially due to educational attainment, "nearly two-thirds of the effect of test scores on earnings is independent of men's education. . . . A 15 point test-score difference between men with the same amount of education is associated with as much as a 14 percent difference in their annual earnings." (4) "The effects of test performance on earnings are not very large relative to the overall earnings gap between the rich and the poor in general." Later, Jencks et al. (1979) noted that the result of their *Who Gets Ahead?* study "suggests that the correlation between adolescent test performance and adult economic success is probably somewhat higher than *Inequality* implied. . . . Our results do not, however, suggest that adult test scores are more closely related to adult economic success than *Inequality* claimed."

These results have been summarized by Seligman (1981) as follows:

> Take, for example, the relationship between I.Q. and income. In its purest form—that is, after subtracting out the effects on income of such factors as age, region of country, and the state of the economy at the time the data were collected—the "coefficient of correlation" is estimated to be quite high, around 0.6. That number signifies a strong and positive, although far from perfect, relationship between I.Q. and income. The square of the number, which is .36, is the so-called coefficient of determination—which tells us that 36% of the variation in income reflects I.Q. differences. Other relationships between I.Q. and income are reported in the second Jencks study. It notes, for example, that among otherwise identical individuals, increasing I.Q. scores by about 15 I.Q. points increases expected lifetime earnings by 20% to 30%. A 15-point difference in brothers' scores is associated with a 13.8% difference in their earnings, assuming that they have the same amount of schooling. So you have to conclude that above-average I.Q.s mean you'll probably have above-average incomes and vice versa.
>
> Similar probabilistic statements might be made about your occupational status. The measurement of status, a major product of the sociology industry, is rooted in surveys in which respondents have ranked many different occupations by the prestige they felt was associated with each. For example, on the famous Duncan Index of Occupational Status, the rankings proceed from the zero given to laborers in tobacco plants to figures in the 90s for, say, judges. While I.Q. is probably the single best predictor of income, educational level is best for occupational status. *Who Gets Ahead?* estimates that high-school graduates outrank elementary-school graduates by 11.6 points and are in turn outranked by college graduates by 25.6 points. (p. 66)

The specific details of the studies reviewed by Crouse and Jencks and

other studies are too technical and numerous to go into in this review, but the general consensus about the role of ability is that its direct effects, especially on income, are consistent but moderate. Most of the effect of ability is in its "indirect" influence on years of education, which then influences attainment. That is, high academic ability allows one to obtain greater amounts of education, which in turn allows one entry to higher-status occupations and thereby to obtain higher incomes. Put another way, high-ability people without a good deal of education are much less likely to have high occupational attainment than high-ability people with a good deal of education. Even moderately able people with many years of education are more likely to have high occupational attainment than high-ability people without many years of education. The mechanism of education—why it should have such effects—is a matter of controversy. The traditional view of educators is that people are taught general competencies that are broadly useful in occupations and, in some cases, specific skills that lead to success in a specific area. Some radical critics, on the other hand, claim that schooling is simply a matter of credentialing and gatekeeping; that is, it is the high school diploma, the college degree, or the professional certification that matters, not the learning of skills that are really essential for the work.

The role of grades tends to be similar to that of academic ability. Higher grades allow individuals to obtain greater amounts of education, which in turn, leads to higher-level occupations and higher income. The direct effect of grades on occupational status and income is fairly small.

Conclusions

In sum, academic ability plays two roles in the process of attaining status and income. The first is a direct effect: the higher one scores on an academic ability test, the higher the attainment. The second role is to increase the probability that the individual will obtain education: the higher one scores on an academic ability test, the more years of education obtained and, subsequently, the higher the level of attainment. The mechanism by which this latter effect takes place is a matter of differing opinions. One view is that the tests are simply surrogates for class-related variables that permeate our educational system. Another is that tests measure the capacity to profit by instruction. Certainly the tests measure abilities and skills necessary to do well and to advance in the educational system. Furthermore, studies of the extent to which grades and academic ability tests merely reflect social class biases indicate that both grades and tests are essentially "class-free."

In sum, there is much evidence that more academically able people are more "successful," in terms of economic and occupational attainment than less academically able people. Many other factors affect success, of course,

so the relationship is not perfect. However, for academically able people to attain success, they have to make use of their ability by attaining education. Without education it appears that raw academic ability will not lead to nearly as high levels of success as with education.

THE TERMAN STUDIES OF THE GIFTED

In 1921, Louis Terman of Stanford University began a study which is still underway. Using the Stanford-Binet Intelligence Test (and other tests in a few cases), he and his coworkers identified a group of 1,528 children, from 3 to 12, most of whom had IQ scores of 140 or above. This group has been followed up intensively for more than 50 years. In the last detailed results (Oden, 1968), the group had reached an average age of 49.5 and a small proportion had died. Thus, the group had clearly had opportunities to demonstrate their capacity for achievement. The reports of the surveys show that the group had attained a very high average level of education: 69% had finished college, 9% had earned doctorates, 8% had obtained law degrees, and 5% held M.D.s. The reports show that their incomes and occupational status were far above the average of the population. Their social contributions were summarized by Oden in 1968:

> In spite of their vocational achievements, the majority of gifted men have found time to participate in civic and community affairs. The most frequent activity for men, reported by 31 per cent, has been participation in youth welfare programs including Boy Scouts, Little League, "Y" recreational activities, Big Brother, and similar groups. Close to 20 per cent have served on school boards, city or county planning commissions, city councils, Grand Juries, boards of directors of philanthropic and welfare organizations, and in various capacities including fund-raising in other community and philanthropic programs. A number of men have won public recognition and honor for their contributions. Among these are 21 men who have received such citations as Citizen of the Year or Man of the Year, Distinguished Civilian Service Award, Distinguished Service to Boyhood medal. At least four men have been appointed at the state level to a Governor's Advisory Board and eight men have served on national advisory committees or councils.
>
> Although many of the men have manifested considerable interest and activity in political as well as civic affairs and community life, the number who have sought election to public office is not very great. One man formerly in the state legislature was later elected to a high office in the executive branch of the state government. Five men have been elected to judgeships—four Superior Court and one Appellate Court. At the local level, at least three men have been elected mayor of their cities. The list of political offices held includes 15 to 20 men who have been elected to county or state central committees of the Republican or Democratic party, as well as several delegates to the national conventions of their party. Others have held office in local Democratic or Republican clubs. Among other political activities

are a hundred or more men who report service as precinct workers, election board officers, and a great deal of miscellaneous party work on behalf of the candidate or party of their choice at election time. In addition to the men who have held elective public office, there are also several who have been unsuccessful candidates for office. These include one who ran for a seat in the United States Senate, one candidate for a Superior Court judgeship, and one who ran for the position of District Attorney. Three men have competed unsuccessfully for election to their state legislature, and several others have been defeated in a try for election to local office.

The most outstanding positions in public service held by the gifted men are appointive. Among these is the head of one of the most important departments, next to cabinet level, in the federal government. Others holding high level appointive positions in the federal government include two ambassadors and five men in executive positions in various divisions of the State Department. Still others are officials in the Federal Reserve Board, Department of Justice, Atomic Energy Commission, National Aeronautical and Space Administration, and Veterans Administration. Three men are assigned to the United Nations in charge of programs in foreign countries and two men are on the staffs of United States senators as special advisers. (pp. 20–21)

Similarly, Oden summarized the writing, professional, and scientific accomplishments of the group:

The men range from top-ranking members of university faculties, famed scientists, men distinguished in the arts and humanities, high level corporate officials and executives, to semiskilled occupations. The group is pretty well concentrated on the upper rungs of the vocational ladder with only a few on the lower steps. There is no evidence that the men with fewer vocational achievements are any less able intellectually than those who have reached high places. In some instances their vocation was determined by educational or occupational opportunities, in others by health, and in still others it was a matter of deliberate choice of a simple, less competitive way of life.

The list of distinctions and honors that have been won is a long one. Three men have been elected to the National Academy of Sciences and two to the American Philosophical Society. Six are included in International *Who's Who,* 46 in *Who's Who in America,* 10 in *The Dictionary of American Scholars,* and 81 in *American Men of Science.* There are many additional listings in regional and other specialized biographical volumes. The achievements of these men also include an impressive number of publications. Some 2500 articles and papers and more than 200 books and monographs in the sciences, arts, and humanities have been published and at least 350 patents granted. Miscellaneous articles (technical, travel, hobby, etc.) number around 350. Other publications include close to 400 short stories, 55 essays and critiques, and a small amount of poetry and several musical compositions. Not included in the foregoing count are the professional output of editors and journalists or the many radio, TV, and motion picture scripts that have been

> authored. Both architects and artists as well as several avocational photographers have had their work chosen for exhibit. In addition to two men, one of whom is a professor and chairman of the art department of a large university and also an artist of considerable distinction, and the other a painter and teacher of private classes in art, 10 men employed in other fields are also gifted painters who devote their leisure time to art. Several of these men, most notably two high school teachers, have produced some distinguished works which have been shown in galleries and won prizes and sales for the artists. Musicians are less frequent than artists among the men: there are, however, three musicians on university faculties, two as heads of the department of music. Four men are performers or choral directors in the field of entertainment. (pp. 19–20)

Although these accomplishments are impressive, unfortunately their significance is difficult to determine. First, one cannot compare these accomplishments with those of any other group. Although the reports meticulously record the percentages and frequencies of the groups' responses to opinion items, their ratings of their marital satisfaction, and so forth, there is no similar detailed information about their accomplishments. In fact, the only information provided about their accomplishments is contained in the paragraphs just quoted. There are no tables, frequencies, percentages, averages, or any numerical information other than the paragraphs. However, even with such information, there would need to be comparable data on the accomplishments of individuals with similar educations and ages but with lower scores on the intelligence test. Without such comparative information, there is simply no way to know whether the highly intelligent Terman group has achieved more than other similar groups which differ only in intelligence scores.

Furthermore, as Oden has suggested, it is difficult to disentangle the role of intelligence from the role of social class in the accomplishments of the group:

> In the total picture, the variables most closely associated with vocational success are a home background in which the parents place a high value on education, encourage independence and initiative, and expect a high level of accomplishment; good mental health and all-round social and emotional adjustment; and the possession of certain traits and characteristics of personality. (p. 92)

However, there are some suggestive comparisons in the Terman study, those of "successful" and "unsuccessful" members of the group. Oden (1968) reported a large number of differences between the two groups denoted as "A" and "C." The A-group members were successful in their professions, their private lives, and their adjustment to life. The typical "A" individual was a productive and lively professional. The C group included people living off estates and doing nothing else, alcoholics, and perennial

students; but more commonly, they were in skilled trades and clerical positions. Although there was a slight difference in childhood intelligence test scores, the largest differences were in several other areas. First, the A parents were of higher social class than the C parents. The A parents had better educations; the A fathers were more often professionals and had more community and professional honors. The A homes had more intellectual resources, such as large libraries. The Cs more often came from broken or divided homes, and homes where money for the children's educational costs was more of an issue than in the A homes. The A parents gave more encouragement to their children's initiative, independence, success in school, and desires to go to college. The As graduated from college in greater numbers than the Cs (92 to 40%); they also achieved many more advanced degrees. Finally, the subjects had been rated by their parents and teachers as children in 1922. The results showed the As to rate higher on "prudence and forethought," "self-confidence," "will power and perseverance," and "desire to excel." In 1940, the subjects were rated again, this time by themselves, their wives, and their parents. The As were rated higher than the Cs on "integration toward goals," "perseverance," "self-confidence," and "absence of inferiority feelings." The subjects rated themselves again in 1950 with the same results, except for the last category. They were also rated by field workers in 1940 and 1950. The variables were selected to cover areas not covered by the earlier ratings. The ratings which best discriminated the groups in 1950 were, in descending order, originality, curiosity, poise, alertness, appearance, attentiveness, attractiveness, and speech.

Perhaps the most striking aspect of these results is the expectedly small role of intelligence scores, compared to the influence of social class, educational level attained, and personality traits reflecting personal stability, social impressiveness, and ambition.

In short, the typical member of the sample who was chosen for his or her scores on the Stanford-Binet in the 1920s has turned out to be a healthy, prosperous middle-class professional who, like most people, is not a genius. In any case, their accomplishments cannot be reliably assessed until there is comparable information about the accomplishments of individuals who obtained the same educations but who had lower intelligence scores. Furthermore, the results of the Terman study only indirectly bear on the question of the overall relationship of academic ability to high-level accomplishment, because the sample is so extreme.

As Keating (1975) has pointed out in an article on possible sampling bias in the Terman study, the sample is even more extreme than 140-plus. If a normal curve of intelligence were used, the mean IQ of a sample above 140 would be 145; Terman's sample averaged 151, a difference significant at the .001 level. The sample includes considerably fewer cases than expected in the

140-to-145 zone and many more than expected in the 146-to-155 zone. The mean of the sample is at a score attained by fewer than 1/10 of 1% of the sample. Thus, even if there were a small correlation between accomplishment and intelligence, a group selected at such an extreme level of intelligence would be expected to show considerable achievements, simply because of the selection ratio (Taylor and Russell, 1939).

In sum, it is difficult to assess the relevance of the Terman study to the question of the relationship between academic ability and accomplishment. The information on their accomplishments is not fully recorded. There is no comparative information on a similar sample of lower IQ. The sample was selected at such an extreme level that generalizations are hazardous. However, it is clear that the persons included in the sample accomplished a good deal; and it is hard to argue that their accomplishments are not due, in large part, to their academic ability.

SUMMARY AND DISCUSSION

What can we conclude about the relationship between academic ability, academic success, and high-level real-life accomplishment? Perhaps the most reasonable position is that academic talent is related to high-level accomplishments in conjunction with several other variables.

The Berkeley studies found essentially no differences between the academic ability or intelligence test scores of the creative and uncreative groups; the major differences between the groups seemed to be personality measures. However, IQs below 120 were seldom found among the groups of scientists, mathematicians, architects, and writers, which suggests that a certain level of academic ability is needed to master these fields.

The Terman longitudinal study shows that individuals with high Stanford-Binet test scores accomplish a good deal, although the level of accomplishment is somewhat uncertain because of the ambiguities of the reports of the project. The comparisons of "successful" and "unsuccessful" members of the sample demonstrated essentially trivial differences in test scores, but showed the importance of personality in this very highly selected group.

The studies of scientists, engineers, and physicians showed scattered correlations between accomplishments and academic ability scores and grades, which may not be surprising considering the diversity of criteria and samples.

The studies of managerial and business success, although sometimes done with sophistication, have to be interpreted cautiously. However, within these samples, there seemed to be low relationships between accomplishments and the individual's estimated ability scores. The studies of particular occupational settings are a mixed group using a wide range of criteria; but, in

general, they suggest some link between academic ability and accomplishments.

The National Merit Scholarship Corporation and the American College Testing Program studies and the other studies of student accomplishment are limited by the fact that they were conducted among college applicants and college students, so the level of accomplishment may not be as high as that in the other studies reviewed. However, there is no reason to believe that the relationship between academic talent and accomplishment should be greatly different for college students than for adults. The accomplishments are real ones, even if they are at the college level. The American College Testing samples represent a broad range of talent. These studies found low positive relationships between academic talent and accomplishment.

The studies of general success and the sociological studies showed some direct effects of academic ability and grades on occupational status and income. Most of the overall effects of ability and grades are due to the greater amounts of education they allow. These results lead to questions about the meaning of education and degrees; but, in general, they suggest that academic ability and academic performance affect academic success and progress, which in turn lead to occupational opportunities.

The meaning of these results may become clearer if we use an analogy from sports. Let us say we have a measure of height, collected at some time in the school years. A student who is tall would be more likely to do well in a sport such as basketball than a shorter student. However, the student would not do at all well without at least some training in basketball. Clearly, excellence in basketball is also influenced by the student's other qualities—coordination, strength, balance, competitiveness, etc. All of this may be considered as analogous to the relationship between measures of academic ability and occupational attainment. Like height in basketball, academic ability is important in occupational attainment; but, like being coached in basketball, education plays a vital role in the attainment process. Again, occupational attainment is also influenced by other factors, such as motivation, inventiveness, special talents, etc. To return to basketball, it is also clear that good coaching can sometimes overcome deficiencies in height. Likewise, good education can help a person with low to moderate academic ability reach high levels of occupational attainment. Also note that the more effective the coaching, the lower the correlation between height and success in basketball, and the more effective the education, the lower the correlation between academic ability and occupational success.

One trend which merits further investigation was found in several of the studies, which indicated that abilities or skills of importance in particular fields are more predictive of success in those fields than tests of general academic ability. For example, in the NSF studies, tests of specific knowl-

edge in the fields of advanced study the applicants planned to pursue predicted later accomplishment better than the more general GRE tests. In several of the industrial studies, tests designed for the professions that were the topic of the study also predicted later accomplishment better than tests of general academic ability. Several other findings were similar, such as the Holland and Richards results showing that, in certain areas, achievement measures did help predict nonacademic accomplishment. For example, a test of English usage and knowledge predicted writing achievement, but a general composite measure did not. Several studies in business showed that second-year business school grades were related to managerial success, but admissions tests were not. All of these results can be interpreted as showing that ability and training which are directly related to a field do predict later accomplishment in the field. Of course, the elements of achievement in many fields are so complex that no test could assess them all. Perhaps this accounts for the consistent result that biographical information about past accomplishments in a given area or about activities similar or preliminary to accomplishment in that area are the best predictors of later accomplishment (Baird, 1976a). Biographical information can assess a broader range of relevant behavior more efficiently than can a test. In any case, general academic ability measures do not have the specificity of these measures, and they are designed to predict general academic success across many different kinds of programs.

Beyond the nature of the tests developed to assess general academic ability, there is another possible explanation for the relatively low relationship between academic ability and accomplishment. Guilford (1968) summarized a considerable number of studies that reported scatterplots of the relationship of scores on "convergent" and "divergent" ability tests which showed a typical shape, as shown in Figure 1.

In Guilford's system "convergent" abilities include academic ability, whereas "divergent" abilities include various capacities that Guilford believes are related to creative accomplishment. Overall, there was a general, small, correlational relationship between convergent and divergent ability tests. Few individuals who were quite low on the convergent ability tests scored high on the divergent ability tests. Although some individuals who scored high on a test of convergent thinking also scored high on tests of divergent thinking, many did not. However, the main point is that the highest scorers on measures of divergent thinking also tended to have high scores on measures of convergent thinking.

Thus, although there may be low overall correlations between convergent and divergent thinking, there may be a critical relationship when the entire spectrum of ability is studied. For example, if we limit the distribution to those who are college applicants scoring (above point A, for example), the

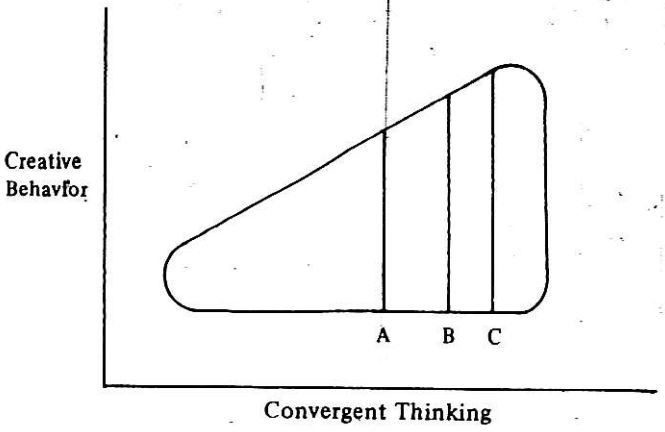

FIG. 1. Fan-shaped distribution of relationship between convergent thinking and creative behavior.

correlation becomes even smaller; and if we limit the distribution to graduates of graduate or professional school (above point C, for example), the correlation would be close to zero. It is clear that, within any of these groups, any further selection on measures of convergent thinking would not increase the proportion of people who demonstrate creative behavior, although the average incidence of creative behavior would be higher than that of groups scoring lower on measures of convergent thinking.

If Guilford's ideas are accurate, it would be virtually impossible to demonstrate a strong relationship between academic ability and creative or high-level accomplishment within any occupational or education group. It may also be difficult to demonstrate a strong relationship across all levels of academic ability, although the highest levels of accomplishment would be expected among people with the highest academic ability. It should be noted that high academic ability is no guarantee of high-level attainment.

This point is similar to that made by observers such as Spaeth (1976), who notes that:

> It should be pointed out that any argument citing a low correlation between educational attainment or test scores and job performance is invalid evidence of the effectiveness of cognitive variables as determinants of occupational status. Since educational attainment provides entry to an occupation and since the incumbent of an occupation is accorded the prestige of that occupation of the day that he enters it, such a low correlation is clearly a matter of false partialling (Gordon, 1968). That is, analyses of job performance must take into account the process by which job incumbents gained entry to their positions. It is all too common for analyses of persons in particular occupations to view the process as if

entering an occupation were not the culmination of years of socialization, training, and selection. This oversight leads to the interpretation of correlations observed within occupations as if they were zero-order correlations pertaining to broader populations.

In sum, it appears that academic ability is clearly a prerequisite to higher levels of education and thus a prerequisite to entrance to various high-level occupations. People who enter these various occupations tend to be similar in a number of other talents and traits as well as in academic ability. Therefore, it is difficult to distinguish among people who tend to share common personal characteristics, educational experiences, and professional values. Although a certain level of academic ability is required for entrance to the training demanded of people in the occupation, it would be very difficult to demonstrate a high correlation between ability and success within these occupations. In fact, considering these difficulties, some of the correlations reported in this review may be surprisingly high. The full force of academic ability is related to educational and occupational attainment, broadly defined. Of course, the same studies show that a host of other variables are also related to attainment, which may tend to mute the direct effect of academic ability. In general, then, academic ability does appear to play a significant role in accomplishment across occupations, while its role within occupations is difficult to demonstrate.

Acknowledgements. This study was conducted while the author was at Educational Testing Service. It was sponsored and funded by the College Board. Material published herein was adapted from College Board Report #ETS RR 83-43 (CBR 82-6) and is used with permission of ETS and the College Board. Points of view or opinions stated do not necessarily represent official College Board position or policy.

REFERENCES

Alexander, K. L., Eckland, B. K., and Griffin, L. J. (1975). The Wisconsin model of socioeconomic achievement: a replication. *American Journal of Sociology* 81: 324–342.

American College Testing Program (1973). *Assessing Students on the Way to College.* Iowa City, Iowa.

Anatasi, A. (1976). *Psychological Testing* (4th ed.). New York: Macmillan.

Andrews, F. M. (1975). Social and psychological factors which influence the creative process. In I. A. Taylor and J. W. Getzels (eds.) *Perspectives in Creativity*, pp. 117–145. Chicago: Aldine.

Astin, A. W. (1964). Socioeconomic factors in the achievements and aspirations of the merit scholar. *Personnel and Guidance Journal* 42: 581–586.

Baird, L. L. (1968). The achievements of bright and average students. *Educational and Psychological Measurement* 28: 891–899.

Baird, L. L. (1969a). The prediction of accomplishment in college: an exploration of the process of achievement. *Journal of Counseling Psychology* 16: 246–254.

Baird, L. L. (1969b). Factors in the continuance of accomplishment from high school to college. *Measurement and Evaluation in Guidance* 2: 5–18.

Baird, L. L. (1972a). Review of Remote Associates Test. In O. K. Buros (ed.), *The Seventh Mental Measurements Yearbook*, Vol. 1, pp. 827–829. Highland Park, N.J.: Gryphon Press.

Baird, L. L. (1972b). Review of Torrance Tests of Creative Thinking. In O. K. Buros (ed.), *The Seventh Mental Measurements Yearbook*, Vol. 1, pp. 836–838. Highland Park, N.J.: Gryphon Press.

Baird, L. L. (1975). The characteristics of medical students and their views of the first year. *Journal of Medical Education* 50: 1092–1099.

Baird, L. L. (1976a). *Using Self-Reports to Predict Student Performance.* New York: College Entrance Examination Board.

Baird, L. L. (1976b). Who goes to graduate school and how do they get there?: The flow of students to advanced education. In J. Katz and R. T. Hartnett (eds.), *Scholars in the Making*, pp. 17–48. Cambridge, MA: Ballinger.

Baird, L. L., and Richards, J. M., Jr. (1968). The effects of selecting college students by various kinds of high school achievements (ACT Research Report No. 23). Iowa City, Iowa: American College Testing Program.

Barron, F. (1965). The psychology of creativity. *New Directions in Psychology II*, pp. 1–134. New York: Holt, Rinehart & Winston.

Barron, F. (1969). *Creative Person and Creative Process.* New York: Holt, Rinehart and Winston.

Bentz, P. G. (1967). Study cited in Campbell et al., *Managerial Behavior, Performance, and Effectiveness.*

Blau, P., and Duncan, O. D. (1967). *The American Occupational Structure.* New York: Wiley.

Block, N. J., and Dworkin, G., eds., (1976). *The IQ Controversy.* New York: Pantheon.

Bowles, S., and Gintis, H. (1976). *Schooling in Capitalistic America: Educational Reform and the Contradictions of Economic Life.* New York: Basic Books.

Burkhart, R. (1967). The relation of intelligence to art ability. In R. L. Mooney and T. A. Razik (eds.), *Explorations in Creativity.* New York: Harper & Row.

Campbell, J. P., Dunnette, M. D., Lawler, E. E., and Weick, K. E. (1970). *Managerial Behavior, Performance, and Effectiveness.* New York: McGraw-Hill.

Cartter, A. M. (1966). *An Assessment of Quality in Graduate Education.* Washington, D.C.: American Council on Education.

Cattell, R. B., and Drevdahl, J. E. (1955). A comparison of the personality profile (16 PF) of eminent researchers with that of eminent teachers and administrators, and of the general population. *British Journal of Psychology* 46: 248–261.

Centra, J. (1974). *Women, Men and the Doctorate.* Princeton, N.J.: Educational Testing Service.

Clark, M. J., and Centra, J. A. (1982). Conditions influencing career accomplishments of Ph.D.s. GRE No. 76-2R. Princeton, N.J.: Educational Testing Service.

Clark, M. J., Hartnett, R. T., and Baird, L. L. (1976). Assessing dimensions of

quality in doctoral education: Technical report of a national study in three fields. Princeton, N.J.: Educational Testing Service.

Clemente, F., and Sturgis, R. B. (1974). Quality of department of doctoral training and research productivity. *Sociology of Education* 47: 287–299.

Cohen, P. A. (1984). College grades and adult achievement: a research synthesis. *Research in Higher Education* 20: 281–294.

Cole, J. R., and Cole, S. (1967). Scientific output and recognition: a study in the operation of the reward system in science. *American Sociological Review* 32: 377–390.

Creager, J. A. (1963). Technical report #22. Washington, D.C.: National Research Council.

Creager, J. A., and Harmon, L. R. (1966). On-the-job validation of selection variables. Washington, D.C.: National Research Council.

Crockenberg, S. B. (1972). Creativity tests: A boon or boondoggle for education? *Review of Educational Research* 42: 27–45.

Crooks, L. A., and Campbell, J. T. (1974). Career progress of MBAs: An exploratory study six years after graduation. PR 74–8. Princeton, N.J.: Educational Testing Service.

Cropley, A. J. (1972). A five-year longitudinal study of the validity of creativity tests. *Developmental Psychology* 6: 119–124.

Cross, P. G., Cattell, R. B., and Butcher, H. J. (1967). The personality patterns of creative artists. *British Journal of Educational Psychology* 37: 292–299.

Crouse, J. (1979). The effects of academic ability. In Jencks (ed.), *Who Gets Ahead?*, chap. 4, pp. 85–121.

Cuca, J. M. Sakakeeny, L. A., and Johnson, D. G. (1976). *The Medical School Admissions Process: A Review of the Literature. 1955–1976.* Washington, D.C.: Association of American Medical Colleges.

Datta, L. (1967). Family religious background and early scientific creativity. *American Sociological Review* 32: 626–637.

Dawes, R. M. (1971). A case study of graduate admissions: application of three principles of human decision making. *American Psychologist* 26: 180–189.

Dellas, M., and Gaier, E. L. (1970). Identification of creativity: the individual. *Psychological Bulletin* 73: 55–73.

Dodd, W. E., Wollowick, H. B., and McNamara, S. J. (1970). Task difficulty as a moderator of long range prediction. *Journal of Applied Psychology* 54: 265–270.

Drew, D. E., and Karpf, R. S. (1975). Evaluating science departments: A new index. Rand Paper Series No. P–5521. Santa Monica, Calif.: Rand Corporation.

Duncan, O. D., Featherman, D. L., and Duncan, B. (1972). *Socioeconomic Background and Achievement.* New York: Seminar Press.

Dunnette, M. D. (1971). The assessment of managerial talent. In P. McReynolds (ed.), *Advances in Psychological Assessment.* Palo Alto, Calif.: Science and Behavior Books.

Edgerton, H. A. (n.d.). Identifying high school seniors talented in science. Mimeo.

Elder, G. H. (1968). Occupational level, achievement motivation and social mobility: a longitudinal analysis. *Journal of Counseling Pyschology* 15: 1–7.

Elton, C. F., and Shevel, L. R. (1969). Who is talented? An analysis of achievement. ACT Research Report No. 31. Iowa City, Iowa: American College Testing Program.

Folger, J. K., Astin, H. S., and Bayer, A. E. (1970). *Human Resources and Higher Education*. New York: Russell Sage.

Gertler, J. M., and Meltzer, A. H. (1970). Selecting creative Ph.D. candidates for admission. *Journal of Experimental Education* 38: 15–18.

Getzels, J. W., and Czikszentimihalyi, M. Problem solving to problem finding. In I. A. Taylor & J. W. Getzels (eds.), *Perspectives in Creativity*. Chicago: Aldine.

Ghiselli, E. E. (1968). Interaction of traits and motivational factors in the determination of the success of managers. *Journal of Applied Psychology* 52: 480–483.

Glass, G. V. (1978). Integrating findings: The meta-analysis of research. In I. S. Shulman (ed.), *Review of Research in Higher Education*, Vol. 5. Itasca, Ill.: F. E. Peacock, 1978.

Golan, S. E. Psychological study of creativity. *Psychological Bulletin* 60: 548–565.

Gough, H. G. (1961). Techniques for identifying the creative research scientist. In D. W. MacKinnon (ed.), *The Creative Person*. Berkeley, Calif.: University of California.

Gough, H. G. (1967). Nonintellectual factors in the selection and evaluation of medical students. *Journal of Medical Education* 42: 642–650.

Gough, H. G. (1976). Studying creativity by means of word association tests. *Journal of Applied Psychology* 61: 348–353.

Graduate Record Examinations (1969). Bulletins of information for candidates. Princeton, N.J.: Graduate Record Examination Board.

Grant, K. L. (1975). Characteristics of successful managers. Paper presented at the American Psychological Association Meetings, Chicago.

Guilford, J. P. (1968). *Intelligence, Creativity, and their Educational Implications*. San Diego, Calif.: Knapp.

Hall, W. B., and MacKinnon, D. W. (1969). Personality inventory correlates of creativity among architects. *Journal of Applied Psychology* 53: 322–326.

Hansen, R. A., and Nevjahr, J. L. (1973). The validity of the Pre-Engineering Ability Test for predicting long term academic success. *Educational and Psychological Measurement* 33: 933–934.

Harrell, M. S., Harrell, T. W., McIntyre, S. H., and Weinberg, C. G. (1977). Predicting compensation among MBA graduates five and ten years after graduation. *Journal of Applied Psychology* 62(5): 636–640.

Harrell, T. W. (1969). The personality of high earnings MBAs in big business. *Personnel Psychology* 22: 457–463.

Harrell, T. W. (1970). The personality of high earning MBAs in small business. *Personnel Psychology* 23: 369–375.

Harrell, T. W. (1972). High earning MBAs. *Personnel Psychology* 25: 523–530.

Harrell, T. W., and Harrell, M. S. (1945). Army General Classification Tests scores for civilian occupations. *Educational and Psychological Measurement* 5: 229–239.

Harrell, T. W., and Harrell, M. S. (1973). The personality of MBAs who reach general management early. *Personnel Psychology* 26: 127–134.

Hays, W. L. (1963). *Statistics for Psychologists.* New York: Holt, Rinehart & Winston.

Helson, R., and Crutchfield, R. S. (1970). Mathematicians: The creative researcher and the average Ph.D. *Journal of Consulting and Clinical Psychology* 34: 250–257(b).

Hemphill, J. K., ed. (1963). *The Engineering Study.* Princeton, N.J.: Educational Testing Service.

Hernstein, R. J. (1973). *I. Q. in the Meritocracy.* Boston: Little, Brown.

Holland, J. L. (1961). Creative and academic performance among talented adolescents. *Journal of Educational Psychology* 52: 136–147.

Holland, J. L., and Astin, A. W. (1962). The prediction of academic, artistic, scientific and social achievement of undergraduates of superior scholastic aptitude. *Journal of Educational Psychology* 53: 132–143.

Holland, J. L., and Baird, L. L. (1968). An interpersonal competency scale. *Educational and Psychological Measurement* 28: 503–510.

Holland, J. L., and Baird, L. L. (1968). The pre-conscious activity scale: The development and validation of a validity measure. *Journal of Creative Behavior* 2: 217–225.

Holland, J. L., and Nichols, R. C. (1964). Prediction of academic and extracurricular achievement in college. *Journal of Educational Psychology* 55: 55–65.

Holland, J. L., and Richards, J. M., Jr. (1965). Academic and nonacademic accomplishment: correlated or uncorrelated? *Journal of Educational Psychology* 56: 165–174.

Holland, J. L., and Richards, J. M., Jr. (1967a). Academic and nonacademic accomplishment in a representative sample of students taking the American College Tests. *College and University* 43: 50–71.

Holland, J. L., and Richards, J. M., Jr. (1967b). The many faces of talent: a reply to Werts. *Journal of Educational Psychology* 58: 205–209.

Howell, M. A. (1966). Personal effectiveness of physicians in a federal health organization. *Journal of Applied Psychology* 50: 451–459.

Howell, M. A., and Vincent, J. W. (1967). The MCAT as related to achievement tests in medicine and to supervisory evaluations of clinical physicians. *Journal of Medical Education* 42: 1037–1044.

Hoyt, D. P. (1966). The relationship between college grades and adult achievement: a review of the literature. *The Educational Record* 47: 70–75. (ACT Research Report 7. Iowa City, Iowa: American College Testing Program, 1965.)

Hudson, L. (1960). Degree class and attainment in scientific research. *British Journal of Psychology* 51: 67–73.

Jacobsen, T. L., Price, P. B., deMik, G. H., and Taylor, C. W. (1965). An exploratory study of predictors of physician performance. (Supported by cooperative Research Project No. OE-3-10-136, Department of Health, Education and Welfare, U.S. Office of Education.)

James, L. R., Ellison, R. L., Fox, D. G., and Taylor, C. S. (1972). Prediction of artistic performance. Paper presented at American Psychological Association. Honolulu, 1972.

Jauch, L. R., and Gleuk, W. F. (1975). Evaluation of university professors research performance. *Management Science* 22: 66–75.

Jencks, C. et al. (1979). *Who gets ahead? The determinants of economic success in America.* New York: Basic Books.

Jencks, C., and Rainwater, L. (1977). The effects of family background, test scores, personality traits and schooling on economic success. Springfield, Va.: National Technical Information Service.

Jones, E. S. (1956). College graduates and their later success. *The University of Buffalo Studies* 22(4).

Jones, F. E. (1964). Predictor variables for creativity in industrial science. *Journal of Applied Psychology* 48: 134–136.

Joseph, A. (1977). *Intelligence, I. Q. and Race—When, How and Why They Become Associated.* San Francisco: R & E Research Associates.

Kamin, L. J. (1974). *The Science and Politics of I.Q.* New York: Lawrence Erlbaum.

Karlins, M., Schuerhoff, C., and Kaplan, M. (1969). Some factors related to architectural creativity in graduating architecture students. *Journal of General Psychology* 81: 203–215.

Kaufman, H. G. (1972). Relations of ability and interest to currency of professional knowledge among engineers. *Journal of Applied Psychology* 56: 495–499.

Keating, D. P. (1975). Possible sampling bias in genetic studies of genius. *Educational and Psychological Measurement* 35: 657–662.

Kelley, L. R., and Goldberg, L. R. (1959). Correlates of later performance and specialization in psychology: a follow-up study of the trainees assessed in the VA selection research project. *Psychological Monographs* 73:(12, Whole No. 482).

Kinloch, G. C., and Perucci, R. (1969). Social origins, academic achievement, and mobility channels: sponsored and context mobility among college graduates. *Social Forces* 48: 36–44.

Korman, A. K. (1968). The prediction of managerial performance: a review. *Personnel Psychology* 21: 295–322.

Kraut, A. I. (1969). Intellectual ability and promotional success among high level managers. *Personnel Psychology* 22: 281–290.

Ladd, E. C., and Lipset, S. M. (1975). *The Divided Academy: Professors and Politics.* New York: McGraw-Hill.

Laurent, V. (1962). Study cited in Campbell et al., *Managerial Behavior, Performance and Effectiveness.*

Lewis, J. (1975). The relationship between academic aptitude and occupational success for a sample of university graduates. *Educational and Psychological Measurement* 35: 465–466.

Locke, E. A. (1963). Some correlates of classroom and out-of-class achievement in gifted science students. *Journal of Educational Psychology,* 54: 238–258.

MacKinnon, D. W. (1962). The nature and nurture of creative talent. *American Psychologist* 17: 484–495.

MacKinnon, D. W. (1968). Selecting students with creative potential. In P. Heist (ed.), *The Creative College Student: An Unmet Challenge.* San Francisco: Jossey-Bass.

MacKinnon, D. W., and Hall, W. G. (1972). Intelligence and creativity. In *Proceedings XVIIth International Congress, International Association of Applied Psychology,* Colloquium 17: The measurement of creativity. Brussels, 1972. EDITEST.

Marston, A. M. (1971). It is time to reconsider the Graduate Record Examinations. *American Psychologist* 26: 653–655.

McCormick, E. J., and Tiffin, J. (1980). *Industrial Psychology* (7th ed.). Englewood Cliffs, N.J.: Prentice Hall.

McNemar, O. (1964). Lost: Our intelligence? Why? *American Psychologist* 19: 871–883.

Mednick, M. T. (1963). Research creativity in psychology graduate students. *Journal of Consulting Psychology* 29: 265–266.

Milgram, R. M., and Milgram, N. A. (1976). Creative thinking and creative performance in Israeli students. *Journal of Educational Psychology* 68: 255–259.

Muchinsky, P. M., and Hoyt, D. P. (1973). Academic grades as a predictor of occupational success among engineering graduates. *Measurement and Evaluation in Guidance* 6: 83–193.

Muchinsky, P. M., and Hoyt, D. P. (1974). Predicting vocational performance of engineers from selected vocational interest, personality, and scholastic aptitude variables. *Journal of Vocational Behavior* 5: 115–123.

Munday, L. A., and Davis, J. C. (1974). Varieties of achievement after college: perspective on the meaning of academic talent (ACT Research Report No. 57). Iowa City, Iowa: American College Testing Program.

Nelson, A. M. (1975). *Undergraduate Academic Achievement as an Indicator of Success.* Washington, D.C.: Civil Service Commission.

Nichols, R. C. (1966). Non-intellective predictors of achievement in college. *Educational and Psychological Measurement* 26: 899–915.

Nichols, R. C., and Holland, J. L. (1964). The selection of high aptitude high school graduates for maximum achievement in college. *Personnel and Guidance Journal* 43: 33–40.

Oden, M. H. (1968). The fulfillment of promise: 40 year follow-up of the Terman gifted group. *Genetic Psychology Monographs* 77: 3–93.

O'Leary, B. S. (1980). College grade point average as an indicator of occupational success: An update. (Personnel Research Report, 80, 23).

Olneck, M. R. (1976). The determinants of educational attainment and adult status among brothers: The Kalamazoo study. Unpublished doctoral dissertation, Harvard University, 1976.

Owens, W. A. (1969). Cognitive, noncognitive and environmental correlates of mechanical ingenuity. *Journal of Applied Psychology* 53: 199–208.

Pallett, J. E., and Hoyt, D. P. (1968). College curriculum and success in general business. *Journal of College Student Personnel* 9: 238–245.

Parloff, M. B., Datta, L., Kleman, M., and Handlon, J. H. (1968). Personality characteristics which differentiate creative adolescents and adults. *Journal of Personality* 36: 528–552.

Perucci, C. C., and Perucci, R. (1970). Social origins, educational contexts, and career mobility. *American Sociological Review* 35: 451–463.

Peterson, O. L., Andrews, L. P., Spain, R. S., and Greenberg, B. G. (1956). An analytic study of North Carolina general practice. *Journal of Medical Education* 31: December, Part II.

Pfeffer, J. (1977). Effects of an MBA and socioeconomic origins on business school graduates' salaries. *Journal of Applied Psychology* 62(6): 698–705.

Porter, A. L., and Wolfe, D. (1975). Utility of the doctoral dissertation. *American Psychologist*, November, pp. 1054–1061.

Price, P. B., Taylor, C. W., Richards, J. M., Jr., and Jacobsen, T. L. (1964). Measurement of physician performance. *Journal of Medical Education* 39: 203–211.

Price, P. B., Taylor, C. W., Nelson, D. W., Lewis, E. G., Loughmiller, G. C., Mathiesen, R., Murray, S. L., and Maxwell, J. G. (1973). Measurement and predictors of physician performance: two decades of intermittently sustained research. (Department of Health, Education, and Welfare Grant No. PM 00017.) Salt Lake City: University of Utah.

Reilly, R. R., and Chao, G. R. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology* 35: 1–62.

Richards, J. M., Jr., Holland, J. L., and Lutz, S. W. (1966). The assessment of student accomplishment in college (ACT Research Report No. 11). Iowa City, Iowa: American College Testing Program.

Richards, J. M., Jr., Holland, J. L., and Lutz, S. W. (1967). Prediction of student accomplishment in college. *Journal of Educational Psychology* 58: 343–355.

Richards, J. M., Jr., and Lutz, S. W. (1968). Predicting student accomplishment in college from the ACT assessment. *Journal of Educational Measurement* 5: 17–29.

Richards, J. M., Jr., Taylor, C. W., Price, P. B., and Jacobsen, T. L. (1965). An investigation of the criterion problem for one group of medical specialities. *Journal of Applied Psychology* 49: 79–90.

Roe, A. (1956). *The Psychology of Occupations*. New York: Wiley.

Roberts, R. J. (1965). Prediction of college performance of superior students. Research Report, Vol. 1, No. 5. Evanston, Ill.: National Merit Scholarship Corporation.

Rowland, K. M., and Scott, W. E., Jr. (1968). Attributes of effective leadership in a formal organization. *Personnel Psychology* 21: 365–377.

Samson, G. R., Grave, M. E., Weinstein, T., and Walberg, H. J. (1984). Academic and occupational performance: a quantitative synthesis. *American Educational Research Journal* 21: 311–322.

Schmidt, H. E. (1973). Personality correlates of the creative architecture student. *Perceptual and Motor Skills* 36: 1030.

Schrader, W. B. (1978). SAT and GRE scores as predictors of career achievement in psychology. Princeton, N.J.: Graduate Record Examinations Board, No. 76–1.

Seligman, P. (1981). Success, luck and careers. *Fortune* 104: 60–72.

Selover, P. R. (1962). Study cited in Campbell et al., *Managerial Behavior, Performance and Effectiveness*.

Sewell, W. H., and Hauser, R. M. (1975). *Education, Occupation, and Earnings*. New York: Academic Press.

Skager, R. K., Shultz, C. B., and Klein, S. P. (1965). Quality and quantity of accomplishments as measures of creativity. *Journal of Educational Psychology* 56: 31–39.

South, J. C. (1974). Early career performance of engineers—its composition and measurement. *Personnel Psychology* 27: 225–243.

Spaeth, J. L. (1976). Cognitive complexity: a dimension underlying the socioeconomic achievement process. In R. M. Hauser and D. L. Featherman (eds.), *Schooling and Achievement in American Society,* pp. 103–131. New York: Academic Press.

Sprecher, T. B. (1954). A study of engineers' criteria for creativity. *Journal of Applied Psychology* 43: 141–148.

Taylor, D. W. Variables related to creativity and productivity among men in two research laboratories. In D. S. Taylor and F. X. Barron (eds.), *Scientific Creativity.* New York: Wiley.

Taylor, D. W., Berry, P. C., and Block, C. H. (1958). Does group participation when using brainstorming facilitate or inhibit creative thinking? *Administration Science Quarterly* 3: 23–37.

Taylor, D. W., and Ellison, R. L. (1975). Moving toward working models in creativity: Utah creativity experiences and insights. In I. A. Taylor and J. W. Getzels (eds.), *Perspectives in creativity,* pp. 191–223. Chicago: Aldine.

Taylor, D. W., Price, P. B., Richards, J. M., Jr., and Jacobsen, T. L. (1965). An investigation of the criterion problem for a group of medical general practitioners. *Journal of Applied Psychology* 49(6): 399–406.

Taylor, H. C., and Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: discussion and tables. *Journal of Applied Psychology* 23: 565–578.

Tenopyr, M. S. (1969). The comparative validity of selected leadership scales relative to success in production management. *Personnel Psychology* 22: 77–85.

Terman, L. M., and Oden, M. H. (1959). *The gifted group at mid-life: Thirty-five years' follow-up of the superior child.* Stanford: Stanford University Press.

Thorndike, R. L., and Hagen, E. (1959). *10,000 careers.* New York: Wiley.

Thurstone, L. L. (1938). *Primary Mental Abilities.* Chicago: University of Chicago Press.

Trattner, M. H., and O'Leary, B. S. (1980). Sample sizes for specified statistical power in testing for differential ability. *Journal of Applied Psychology* 65: 127–134.

Wallach, M. A., and Kogan, N. (1969). *Modes of Thinking in Young Children: A Study of the Creativity-Intelligence Distinction.* New York: Holt, Rinehart & Winston.

Wallach, M. A., and Wing, C. W., Jr. (1969). *The Talented Student: a Validation of the Creativity-Intelligence Distinction.* New York: Holt, Rinehart and Winston.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corporation.

Weinstein, A. G., and Srinivasau, V. (1974). Predicting managerial success of Master of Business Administration (MBA) graduates. *Journal of Applied Psychology* 59: 207–212.

Weitzman, R. A. (1972). It is time to re-consider the GRE: A reply to Marston. *American Psychologist* 27: 236–238.

Werts, C. E. (1967). The many faces of intelligence. *Journal of Educational Psychology* 58: 198–204.

Wing, C. W., Jr., and Wallach, M. A. (1971). *College Admissions and the Psychology of Talent.* New York: Holt, Rinehart & Winston.

Wingard, J. R., and Williamson, J. W. (1973). Grades as predictors of physicians' career performances: an evaluative literature review. *Journal of Medical Education* 48: 311–322.

Wise, D. A. (1975). Academic achievement and job performance. *American Economic Review* 65(3): 350–366.

Wolfson, K. P. (1976). Career development patterns of college women. *Journal of Counseling Psychology* 23: 119–125.

Received April 4, 1985