

Barrett and Depinet (1991) claimed that Boyatzis (1982) had said that "competency testing was distinct from, and superior to, assessment centers" (p. 1021). The actual quotation was, "Even assessment centers, as they are often designed, currently do not assess competencies as much as they assess performance of job functions. . . . Some organizations have designed their assessment centers and selection techniques to assess competencies" (Boyatzis, 1982, p. 250).

To be clear, assessment centers are a method for organizing, administering, and typically, returning data to participants. Assessment centers can be, and are, used to assess competencies, often incorporating the same operant exercises, such as the leaderless group discussion or CII. Even with the relatively high "efficiency" with which the CII can be used to assess competencies, the desirability of multimethod assessment, multievent sampling, and comprehensiveness of measures dictates the inclusion of the CII with other measures in research and applications.

Concluding Comment

Although dialogue and debate are the lifeblood of scholarship, important points can get lost in the cross fire of dueling references. In the 1973 article, McClelland concluded with six principles for understanding human capability while respecting the integrity of research methods and individuality. He claimed that tests should use criterion sampling, be designed to reflect changes in what the individual has learned, make public how to improve on the characteristics tested, assess competencies involved in clusters of life outcomes, involve operant as well as respondent behavior, and sample operant thought patterns.

REFERENCES

- Barrett, G. V., & Depinet, R. L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist*, 46, 1012-1024.
- Boyatzis, R. E. (1982). *The competent manager: A model for effective performance*. New York: Wiley.
- Boyatzis, R. E. (in press). Developing the whole student: An MBA required course called Managerial Assessment and Development. *Journal of Management Education*.
- Campbell, J. P., Dunnette, M. D., Lawler, E. E., III, & Weick, K. E., Jr. (1970). *Managerial behavior, performance, and effectiveness*. New York: McGraw-Hill.
- Crooks, L. A., Campbell, J. T., & Rock, D. A. (1979). *Predicting career progress of graduate students in management* (Report prepared by the Educational Testing Service for the Graduate Management Admissions Council).

- Princeton, NJ: Educational Testing Service.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327-358.
- Howard, A., & Bray, D. W. (1988). *Managerial lives in transition: Advancing age and changing times*. New York: Guilford Press.
- McClelland, D. C. (1973). Testing for competence rather than for "intelligence." *American Psychologist*, 28, 1-14.
- Mentkowsky, M., O'Brien, K., McEachern, W., & Fowler, D. (1982). *Developing a professional competency model for management education* (Final report to the National Institute of Education: Research Rep. No. 10). Milwaukee, WI: Alverno College.
- Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N., Werner, S., Burnett, J. R., & Vaughan, M. J. (1992). Studies of the structured behavioral interview. *Journal of Applied Psychology*, 77, 571-587.
- Ronan, W. W., & Latham, G. P. (1974). The reliability and validity of the critical incident technique: A closer look. *Studies in Personnel Psychology*, 6, 53-64.
- Thornton, G. C., III, & Byham, W. C. (1982). *Assessment centers and managerial performance*. San Diego, CA: Academic Press.

The Knowledge-Testing-Educational Complex Strikes Back

David C. McClelland
Department of Psychology,
Boston University

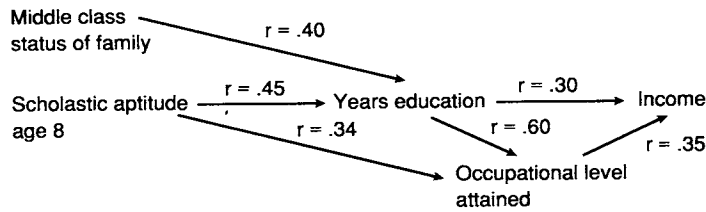
Barrett and Depinet (1991, October) were distressed because an article I once wrote on "Testing for Competence Rather Than for 'Intelligence'" (McClelland, 1973) has been widely influential, despite the fact that its assertions are wrong, "contradicted by other evidence." From my point of view, they have distorted what I said so that it can be more easily attacked, but to attempt to reply to them point by point seems unproductive because there is not the space to do it properly and because the reader is unlikely to read my original article to compare what each of us says it said. But I simply cannot resist examining more closely one of the examples they cited to show how wrong I was. They were delighted to discover and highlight in a table an early study (Nicholson, 1915) that showed that grades in college were related to success in life, as reflected in inclusion in *Who's Who*. Because I had contended that academic performance (as opposed to academic credentials) had little relation to life success, they must have chuckled at my expense, inasmuch as the college involved was Wesleyan, where I went and taught and used to cite cases of A students who had not amounted to much. But what they failed to realize (or point out)

was that at that time nearly all professors in the top five New England colleges were in *Who's Who*, and 55% of the men included in Dean Nicholson's sample were professors! So, top scholastic performance had prepared people to be good professors so that they could get into *Who's Who*, thus illustrating my point, not theirs. For good professors continue to do well the things they did as good students (writing papers, etc.). So, good grades should predict success as a professor because they sample criterion behavior—which is just what I had said tests should do.

But, rather than engage in such gamesmanship, it seems better to start over by laying out some of the relationships supposedly under dispute from a recent follow-up study we have been making of children whose parents' child-rearing practices were studied by Sears, Maccoby, and Levin (1957). The families of origin were classified according to social class (blue collar vs. white collar) in the standard way used by sociologists (a combination of education and occupational level of parents), and most of the children were given a scholastic aptitude test at age 8 (Stanford Achievement Test Form J2, 1931). In later follow-ups at age 31 and age 41, information was obtained on years of education, occupational level attained, income, and other life outcomes. The correlations among these measures appear in Figure 1.

All of the correlations are significant at $p < .01$ as sample size varies from 60 to 85. Barrett and Depinet may well be surprised that I would display such figures, because they might seem to "admit" everything they were accusing me of being wrong about. For instance, scholastic aptitude is significantly correlated with occupational level attained, an important life outcome variable. I was not questioning this fact but the *interpretation* testers often give this correlation—namely, that higher scholastic aptitude *causes* or is directly responsible for being able to perform well in higher level demanding occupations. I claimed and still claim (and they seem even to agree) that this correlation is mediated by the close connection of academic aptitude to more schooling ($r = .45$). As I argued previously, rather vehemently, the one thing scholastic aptitude tests *do* predict without question is school performance—as they are in effect sampling the very kinds of behavior required in school. And there is an even closer tie of more schooling to entry into higher level occupations ($r = .60$), because schooling serves a credentialing function for the professions. Thus, if you partial out the effect of education, the relation of scholastic aptitude to life outcomes (e.g., occupational level, income) is effectively reduced to zero.

Figure 1
Correlations Among Measures



That *may* be interpreted to mean that ability as measured in these tests has nothing *directly* to do with professional performance but only with getting through the educational hurdles necessary to get the educational credentials needed. Education seems to be the key determinant of success: Many have shown that, granted equivalent levels of education, academic aptitude test scores do not relate to measures of vocational success—even among Terman’s gifted children (Cronbach, 1984). And ability tests have been one of the key determinants of who gets an education.

The same response may be given to Barret and Depinet’s (1991) repetition of the well-known fact, which again I accepted, that knowledge tests do relate to supervisor’s ratings in various jobs. What I was concerned about was not whether this fact existed but how testers *interpreted* it. So, I checked one of the best of the recent references they provided in this area (Schmidt, Hunter, & Caplan, 1981) to see if the authors had paid any attention to my complaints that caution is seldom observed in *interpreting* these correlations. Despite reporting on hundreds of correlations and thousands of subjects, using the most sophisticated statistical methods, so far as I could determine, they did not once mention or correct for educational level. Generally, they found that *any* knowledge test (chemical, mechanical, or arithmetical) correlated at about .20 with supervisors’ ratings of job performance as operators or maintenance people, leaving the definite impression that knowledge or ability (of a surprisingly diverse kind!) is what made them better performers. Yet, it could easily be that education mediates the relation of ability to performance as in the above matrix. My main gripe now as it was then is that they didn’t even bother to look at this relationship. So, simply by omission, they left the unsuspecting reader to believe they have *proved* that knowledge of chemistry, or whatever, is what makes a better maintenance person.

And what about the relation of schooling to a life outcome such as income? The

relation is significant ($r = .30$) in the matrix above. But again, it is clearly mediated by the contribution of schooling to occupational level: You cannot be a high-earning professional without the credential provided by higher education. So, if you partial out occupational level, the relation of education to income is reduced to insignificance. But what about doing better in school? Here our difference is one of degree. I said “performance within grade was related only slightly” to life success. Path analyses of variables contributing to later life success (Whitla, 1975) in the fields of law, medicine, teaching, and business for Harvard graduates have shown that rank list or graduating with honors yields path coefficients around .20 to .22 with later success (whereas MAT and SAT scores have zero order or negative coefficients). If they want to consider those relations “substantial” rather than slight, it is all right with me, but whatever you call them, it is obvious that many competencies other than academic talent are involved in life success in these areas.

Another and rather independent contributor to schooling, as I argued previously, is social class or family background ($r = .40$, in the matrix above). Class, as measured in the standard way (by family occupation and education levels), correlates also with grades in school (White, 1982) at an average level of .325, a relationship that Barrett and Depinet (1991) dismissed as arguing that it is a “misperception” that social class contributes to educational attainment. Normally, social class also contributes directly to scholastic aptitude test scores, but in the sample for the matrix above, the correlation is only .11, because it was drawn from fairly affluent Boston suburbs, in which there were no really disadvantaged families. Where poor families are included, the correlations run from .27 to .35 (Havighurst, Bowman, Liddle, Matthews, & Pierce, 1962, as cited in McClelland, 1973; Valencia, Henderson, & Rankin, 1985). Recent work by sociologists such as Jencks (1977) have shown that family advantage contributes strongly to occu-

pational success, even when aptitude and educational levels are controlled for.

The most extraordinary claim Barrett and Depinet (1991) made is that my assertion that this whole testing set-up is unfair to underprivileged minorities “has been refuted by scientific evidence.” The only evidence they adduce is that ability test scores predict academic performance as well for Blacks as for Whites, but whoever doubted that? I never did, inasmuch as I regard such tests as criterion sampling for what goes on in schools. The fairness issue was totally different. Put yourself in the position of a poor African American, meaning that your family has neither the attitudes nor financial resources to keep you in school, because of the costs of staying in school, which poor people cannot afford. You are also likely to have a lower scholastic aptitude test score and poorer grades because of your disadvantaged family background (Klitgaard, 1985). These test scores and lower grades regularly discourage students from staying in school, and as we have seen, it is access to schooling that, overall, contributes most to getting ahead. Furthermore, SATs and MATs have been used to screen Black applicants out of better colleges, which provide improved access to better jobs. That may not be unfair if doing well *in college* is the criterion, but it is doing well in life we should be concerned with, and the ability-testing-education establishment screens disadvantaged people out of school, so that they have no opportunity to show that they can do well at various jobs because of other competencies they may have.

This brings up the question of competency testing, which Barrett and Depinet (1991) dismissed as a very poor and untested alternative to ability testing. I would have to agree that the competency *testing* alternative has not been a success so far, and I can give examples that I think will make it clear what the reasons are for this failure.

In 1973, McBer and Co., for which I was a consultant, was asked by the Massachusetts Civil Service Commission to help them solve a problem they had run into. A number of human service workers (HSWs) in the third largest state-employee category (case or employment aides, special or social service assistants, etc.) were being paid with federal funds designed to support such work by disadvantaged people. For the most part, they were middle-aged African-American women with about a sixth-grade education, whose supervisors reported they were doing good work. So, their departments wanted to transfer them to regular state-supported positions; yet, they could not do so under the rules of the State Civil Service Commission because they could not pass the General Aptitude Test Battery (GATB) at a high

enough level. In this case, there was direct evidence that bypassing the ability test screen enabled these people to show they could be successful in these occupations. But, note that the GATB screening was keeping them out of such jobs and that such discrimination provides empirical support for Barrett and Depinet's contention that poor ability is associated with lower job success. Yet, using such tests prevents psychologists from collecting data on how well people do with lower GATB scores.

This bothered us (McClelland & Fiske, 1974) and the Civil Service Commission. They wanted us to prepare a set of competency-based tests that would be valid enough to demonstrate that these people were capable of performing at a high level. After carrying out job analyses and behavioral event interviews (the nature of which Barrett and Depinet, 1991, misrepresented) of more and less successful HSWs, we designed criterion-based tests that we validated against existing HSWs who had been nominated as "outstanding" or "typical" by supervisors. The criterion measure of outstanding performance did not correlate significantly with race, sex, age, or education. To simplify greatly, we determined that HSWs needed two types of competencies to perform well in these positions: They must be able to relate to people and they must be able to perform certain clerical tasks accurately, such as processing benefits.

We (McClelland & Fiske, 1974) included, therefore, a clerical aptitude-type test, and like nearly all such tests it correlated positively with education and negatively with being Black. Overall, it correlated $-.08$ with the criterion, but we argued that it tested a threshold competency, that is, that there was obviously some minimum level of being able to read, write, and figure necessary to do the paperwork required on these jobs. So, we simply set as a passing score the lowest number correct obtained by any of the HSWs classified as outstanding. If I would change anything I wrote in the 1973 article, it would be to explain more fully the need for establishing *threshold* competencies, the way we did here, which can be done by traditional ability tests. Most of our energy went into developing a test of the human relations skills needed for the job. As a measure, we developed a Scenarios Test, which consisted of actual interviews of clients tape-recorded in the field, together with a still photograph of the clients (today, we would use videotapes). The testees listened to the tapes (not always easy to understand because of background noise and the use of "street talk") and answered factual questions about what they had heard, wrote out what they would recommend in the case (scored against expert judgment),

answered questions as to what the attitudes of the clients were on various issues (checked against what the clients actually said), and so forth. In other words, this was sampling criterion behavior. This is what HSWs often do—interview, review facts and attitudes, and make inferences and recommendations. Three such measures taken together correlated significantly with the criterion, and we identified a cutting score on the test battery that would ensure that most of the people at that score or above would be classified as "outstanding." Furthermore, scores on the test battery were not significantly associated with sex, race, or education. Thus, the Massachusetts Civil Service Commission felt justified in using this validated instrument as a screening alternative to the GATB, which incidentally had never been validated against superior performance on the job! So a number of women who were being screened out by the ability-testing-educational complex were enabled to get permanent positions and continue to do what they were doing well by the use of validated competency tests.

So, why hasn't more of this been done? We at McBer were naive enough to think that now that we had built a validated test, the Civil Service Commission would use it more generally. What happened soon set us straight: The social workers, a number of them with master's degrees, were very upset by the possibility that such a test might be substituted for the GATB and threatened a job action if the Commission used it. Note that the new competency tests did not correlate with years of education! That meant quite simply to some social workers that people might get the jobs who had not gone to social work school and that master's degrees might turn out to be worthless, despite the time, energy, and opportunity costs put into them! No way! It was then that we began to understand how tightly the achievement (knowledge) testers are tied in to the educational establishment. The educators are not threatened by the GATB, because people with college degrees would do well on it or they could not have gotten into or through college.

Because, doubtless, in their reply to my reply, Barrett and Depinet will want to show that I am still wrong, that it is not necessary to do partial correlations, and so on, I would like to ask them to respond directly to the situation of these poorly educated Black women. What would they have done, if anything?

Or to consider another example, the U.S. Information Agency (part of the Foreign Service) came to McBer in the 1970s because they wanted very much to have minorities representing this country abroad. However, they could pick only about 400 of

the highest scorers on ability and knowledge tests provided by the Educational Testing Service (ETS) to come in for an interview, out of 2,400 taking the tests. And they found very few minority candidates who scored in the top 17% of the scoring distributions. They also knew that some Blacks recruited laterally into the service were performing very well on the job, even though they were not in the top 17% on the ability-knowledge tests. So we (McClelland & Dailey, 1973) carried out behavioral event interviews on outstanding and average agency representatives in various countries and designed tests assessing the competencies shown more often by the outstanding people as they carried out their work (criterion sampling). We demonstrated that these specially designed competency tests (including measures of speed of learning political networks, correct judgment calls in critical episodes, and social sensitivity) were valid as a battery in discriminating significantly between outstanding and average agency representatives. The competency test battery also did not discriminate against Blacks or women and screened in different people from those chosen by the traditional ability-knowledge tests provided by the ETS. That was as far as we got: We did not have the political or financial clout to compete with ETS to get the new competency tests used by the agency. Note that we had proved that the competency tests were valid and nondiscriminatory, whereas the ETS tests had never been validated against success on the job and were known to be discriminatory.

To understand why we got nowhere, consider the test we used of social sensitivity, which was drawn from the Rosenthal (1979) Test of Non-Verbal Sensitivity. Outstanding representatives of the United States in the Foreign Service scored higher than average representatives in understanding content-filtered speech as presented on a tape in two independent samples. In other words, they were better at understanding the feelings behind words when the literal content was unavailable to them. No matter how valid or "reasonable" such a competence measure might be, it would not be acceptable to the knowledge-testing-educational establishment because (a) it does not have some of the structural characteristics knowledge tests have, and (b) social sensitivity is not something Ivy League colleges, from which U.S. diplomats are disproportionately drawn, are likely to regard as the product of the higher education they provide. On the other hand, graduates of the same colleges who were judging the worthwhileness of our competency tests could readily agree that elegance of English expression, as assessed by ETS, is some-

thing their colleges taught and something that ought to characterize better diplomats. Yet, we could find little evidence that better diplomats wrote reports in better English that made a real difference on the job. How would Barrett and Depinet deal with this problem? So far as I know, the knowledge-ability tests are still being used to screen out minority candidates, unfairly because of their handicapped backgrounds, who could do the job perfectly well, as shown by valid competency tests, when there is no evidence that ability test scores at the very high level required at screening are related to superior performance on the job.

I believe that well-designed competency-based tests could make an important contribution to selecting people who are better suited for various jobs, but to judge by their inability to influence the knowledge testing establishment to date, they will not be developed until there is a strong commitment by psychologists to develop them, together with the financial support necessary to do the job right. The Educational Testing Service has turned out to be a handicap in this effort. It was set up to serve the educational establishment, and that it has done very well. But its resources and prestige are so great that it has moved into fields in which its tests are not applicable, and their unvalidated use actually prevents the development of more appropriate instruments. Barrett and Depinet seem determined to maintain the knowledge testing status quo by damning the whole enterprise of trying to develop alternative types of tests.

Competency assessment through tests has not gone anywhere because the ability-testing-education establishment has a firm hold on how people are allowed to get ahead in this country, as I said in 1973. Competency assessment through interviews, on the other hand, has been more of a success because (a) interviews are more acceptable in many contexts, particularly in business, and (b) interviews provide concrete behavioral information that can be used to design training courses to develop those competencies (McBer, 1987). There is ample evidence that competencies identified through interviews for high-level occupations have led to better selection and training in business. Unfortunately, not many of these studies have been published, in part because some of the information is proprietary and in part because the psychologists doing them must spend their time earning their salaries rather than writing papers for publication. But two recent books, at least, give an introduction to how such studies are done for those who want to understand the approach better (Nygren & Ukeritis, 1993; Spencer & Spencer, 1993).

REFERENCES

- Barrett, G. V., & Depinet, R. L. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist*, 46, 1012-1024.
- Cronbach, L. J. (1984). *Essentials of psychological testing* (4th ed.). New York: Harper & Row.
- Havighurst, R. J., Bowman, P. H., Liddle, G. P., Matthews, C. V., & Pierce, J. V. (1962). *Growing up in River City*. New York: Wiley.
- Jencks, C. (1977). *Who gets ahead? The determinants of economic success in America*. New York: Basic Books.
- Klitgaard, R. (1985). *Choosing elites*. New York: Basic Books.
- McBer staff. (1987). *A history of the U.S. Navy leadership and management education and training program*. Boston: McBer & Co.
- McClelland, D. C. (1973). Testing for competence rather than for "intelligence." *American Psychologist*, 28, 1-14.
- McClelland, D. C., & Dailey, C. (1973). *Evaluating new methods of measuring the qualities needed in superior Foreign Service officers*. Boston: McBer & Co.
- McClelland, D. C., & Fiske, S. T. (1974). *Report to the executive officer of manpower affairs on validation of a human service worker test*. Boston: McBer & Co.
- Nicholson, F. W. (1915). Success in college and after life. *School and Society*, 12, 229-232.
- Nygren, D. J., & Ukeritis, M. D. (1993). *The future of religious orders in the United States: Transformation and commitment*. Westport, CT: Praeger.
- Rosenthal, R. (Ed.). (1979). *Skill in non-verbal communication*. Cambridge, MA: Oelgeschlager.
- Schmidt, F. L., Hunter, J. E., & Caplan, J. R. (1981). Validity generalization results for two job groups in the petroleum industry. *Journal of Applied Psychology*, 66, 261-273.
- Sears, R. R., Maccoby, E. E., & Levin, H. (1957). *Patterns of child-rearing*. Evanston, IL: Row Peterson.
- Spencer, L. M., & Spencer, S. M. (1993). *Competence at work*. New York: Wiley.
- Valencia, R. R., Henderson, R. W., & Rankin, R. J. (1985). Family status, family constellation, and home environmental variables as predictors of cognitive performance of Mexican-American children. *Journal of Educational Psychology*, 77, 323-331.
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, 91, 461-481.
- Whitla, D. K. (1975). *Value added: Measuring the impact of undergraduate education*. Cambridge, MA: Office of Instructional Research and Evaluation, Harvard University.

Empirical Data Say It All

Gerald V. Barrett
 Department of Psychology,
 University of Akron

After considering Boyatzis's (1994, this issue) and McClelland's (1994, this issue)

comments and reviewing additional reports provided by them, the conclusion expressed by Barrett and Depinet (1991, October) has been further reinforced. "If McClelland's concept of competencies is to make a contribution to the field of psychology, he must present empirical data to support his contention" (p. 1021). This point is illustrated by three sets of data presented below.

The first set of data is from a validation study of U.S. Foreign Service Information Officers (FSIOs) conducted by McClelland and Dailey (1973). The U.S. Knowledge Test was obviously the most reliable and valid test of the test battery. Other tests, even when rescored on the validation same and *not* cross validated, had unacceptable reliabilities and validities. The Learning Social Relations (Discrepancies) test had a validity of .20, but McClelland and Dailey did not recommend its use because "it seemed worrisome to think that a person who was really wrong in overestimating what he would do would be especially qualified as an F.S.I.O." (p. 9).

The Judging Critical Incidents Test was described as the "most job-related material to be used in any of the tests" (McClelland & Dailey, 1972, pp. 23-24). Even when rescored, its reliability and validity were below professional standards. (See Table 1.)

The second set of data was from a validation study of human service workers conducted by McClelland and Fiske (1974). Of 18 correlations, which were not rescored or "age corrected," only one was significant. (See Table 2.)

The highest correlation of more than .35 computed in the two validation studies conducted by McClelland and his associates was from a knowledge test. This result is consistent with a meta-analysis involving over 500 coefficients and over 350,000 individuals, showing that validities were nearly twice as high for job-specific knowledge tests compared with off-the-shelf tests (Dye, Reck, & McDaniel, 1993).

The third set of data illustrates job sample tests that were both content valid and demonstrated criterion-related validity (Hall & Barrett, 1977). These were *not* paper-and-pencil tests. These tests were developed for individuals who had limited education and who were developmentally challenged. The reliabilities and validities were much higher than the two previous data sets from McClelland and his associates. This illustrates that nontraditional reliable and valid tests can be constructed.

McClelland (1994) asks what we would do in human services and foreign service situations. First, I would develop a separate test for each of the six jobs. McClelland and Fiske's (1974) job description clearly