

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

Table 1*Tests for U.S. Foreign Service Information Officers*

Test	Reliability (Cronbach's α)	Validity
Empathy	.20	.11
Preference (variety)	—	-.02
Self-description		
Management skills	—	-.13
Empathy	—	-.05
Political sense	—	-.06
Promotion	—	-.01
Patient resilience	—	.12
Learning social relations		
Discrepancies	—	.20
Accuracy	—	.07
Sum	—	-.08
Creativity in problem solving	—	.08
Judging critical incidents ^a	.36	.16

Test	Reliability (split half)	Validity
Learning social relations (Correct on Problem B - Correct on Problem C) ^b	.80	.21
Background questionnaire		
Relevant experience (participation in student government in high school plus college academic record minus sum of participation in musical activities in high school and college)	—	.16
Job preparation index (high school teaching minus sum of public relations and writing job experience)	—	.13
U.S. Knowledge Test	.70	.32

Note. $N = 115$. Data are from *Evaluating New Methods of Measuring the Qualities Needed in Superior Foreign Service Officers* by D. C. McClelland and C. Dailey, 1973, Boston: McBer.

^a10% of answers changed to correspond with answers given by superior Foreign Service Information Officers; original validity coefficient not reported.

^bRescored because sample of better Foreign Service Information Officers did worse on Problem C, contrary to original prediction.

showed little overlap in job duties between case aide and senior interviewer, for example. Second, I would use a content-oriented test construction approach for each separate job, as was done in the Hall and Barrett (1977) study. (See Table 3.)

Third, I would construct reliable tests. Fourth, if a criterion-related validity study had been conducted, I would not rescore

tests and report results unless they had been cross validated. These rescored tests could not be included in a battery. Fifth, I would not age-correct scores as this would violate Civil Service rules, professional guidelines, and the Civil Rights Act of 1991. Sixth, I would use recognized professional terminology. If a content valid job sample approach is being advocated, it should not be referred to as a measure of "competencies." In the early 1900s, Münsterberg (1913), at Harvard, developed and validated job sample tests. Job sample tests are not a new idea. In the same vein, if the critical incident technique was used by Boyatzis, I see no value in labeling it the Behavioral Event Interview.

Seventh, I would not confuse observed or reported critical incidents with a test. Critical incidents may help in constructing a test, but they are not tests themselves. This is a mistake Boyatzis (1994) continues to make. He has reported no empirical validation data, even in his most recent research. The "Judging Critical Incidents Test" was neither reliable nor valid, even when rescored. Eighth, I would use recognized constructs when I developed tests (Barrett, 1992). *Competencies* is a confusing and ambiguous concept (Sparrow & Bognanno, 1993). For instance, what "competency" is being measured by McClelland's Relevant Experience Test, which is "participation in high school student government plus col-

Table 2*Tests for Human Service Workers (Case Aides, Social Service Technicians, Employment Aide/Agent, Senior Interviewers, Attendant Nurses, and Special Service Assistants)*

Test	Reliability (split half)	Validity
Clerical tests		
Bus schedules	—	-.08
Telephone test	—	ns
Scenario tests		
Positive bias	.76	.29
Unrealistic optimism	—	.11
Casework ^a	.46	.28
Diagnostic listening test ^b	.31	.25
Learning social relations (5 scores)	—	ns
Attitudes toward human service work (5 scores)	—	ns
Job frustration (2 scores)	—	ns

Note. $N = 53$. Data are from *Report to the Executive Officers of Manpower Affairs on Validation of a Human Service Worker Test* by D. C. McClelland and S. T. Fiske, 1974, Boston, McBer.

^aRescored on the basis of results from the validation sample.

^bAge corrected.

Table 3*Job Sample Tests for Housekeeping and Laundry Jobs*

Job/Content valid test	Reliability (test-retest)	Validity (supervisor ratings)
Housekeeping/ Sweeping, mopping, dusting, window cleaning, and trash pickup	.92	.65
N	65	54
Laundry/Sort by color, type, and identification marks; fold; and load into bags)	.83	.86
N	30	22

Note. Data are from "Payday for Patients: Federal Guidelines or a Job-Sample Approach?" by R. Hall and G. V. Barrett, 1977, *American Psychologist*, 32, p. 587. Copyright 1977 by the American Psychological Association.

lege academic record minus musical activities in high school and college"? Ninth, I would not set cutoff scores of "threshold competency" using a test that has no validity. Professionally acceptable, legal approaches are available for setting cutoff scores (Cascio, Alexander, & Barrett, 1988). Tenth, I would not claim that my test batteries were valid and nondiscriminatory and tests by other organizations (i.e., Educational Testing Service) were not valid and discriminatory without empirical evidence. In my review of McClelland's validation studies, I would conclude that his tests were not valid and that there was no professionally acceptable evidence of their being nondiscriminatory.

McClelland (1994) asserts that we believe only in traditional standardized tests. I recently reviewed the results from 17 validation studies (Barrett, 1993), in which specific information processing tests were computer administered using time and errors as scores. These nontraditional tests had equal or higher validities than traditional cognition ability tests, with less adverse impact. I support the use of these reliable and valid nontraditional tests and this approach to test development.

McClelland (1994) chides the field of industrial/organizational psychology for not including educational level in validation studies. Most studies do in fact include educational level, but usually most individuals in any given sample have a very similar educational experience, which prevents educational level from being related to performance because of the restricted variance on this variable. This was true with the 17 studies mentioned above (Barrett, 1993).

McClelland (1994) admits that competency testing has not been a success. He attributes this to other forces such as the "knowledge-testing-educational complex." A simpler explanation is that the approach has not resulted in professionally acceptable levels of either reliability or validity, as illustrated by the first two tables.

REFERENCES

- Barrett, G. V. (1992). Clarifying construct validity: Definitions, processes, and models. *Human Performance*, 5, 13-58.
- Barrett, G. V. (1993, May). *The bridge between professional practice and research in industrial/organizational psychology*. Invited address at the 8th annual meeting of the Society for Industrial and Organizational Psychologists, San Francisco.
- Barrett, G. V., & Depinet, R. (1991). A reconsideration of testing for competence rather than for intelligence. *American Psychologist*, 46, 1012-1024.
- Boyatzis, R. E. (1994). Rendering unto competence the things that are competent. *American Psychologist*, 49, xx-xx.
- Cascio, W. F., Alexander, R. A., & Barrett, G. V. (1988). Setting cutoff scores: Legal, psychometric, and professional issues and guidelines. *Personnel Psychology*, 41, 1-24.
- Dye, D. A., Reck, M., & McDaniel, M. A. (1993). The validity of job knowledge measures. *International Journal of Selection and Assessment*, 1, 153-157.
- Hall, R., & Barrett, G. V. (1977). Payday for patients: Federal guidelines or a job-sample approach? *American Psychologist*, 32, 586-588.
- McClelland, D. C. (1994). The knowledge-testing-educational complex strikes back. *American Psychologist*, 49, xx-xx.
- McClelland, D. C., & Dailey, C. (1972). *Improving officer selection for the Foreign Service*. Boston: McBer.
- McClelland, D. C., & Dailey, C. (1973). *Evaluating new methods of measuring the qualities needed in superior Foreign Service officers*. Boston: McBer.
- McClelland, D. C., & Fiske, S. T. (1974). *Report to the executive officers of manpower affairs on validation of a human service worker test*. Boston: McBer.
- Münsterberg, H. (1913). *Psychology and industrial efficiency*. New York: Houghton Mifflin.
- Sparrow, P. R., & Bognanno, M. (1993). Competency requirement forecasting: Issues for international selection and assessment. *International Journal of Selection and Assessment*, 1, 50-58.
- In the September 1992 issue, Charles Kiesler documented how mental health policy is doomed to fail if it continues to follow the lead of a flawed national health policy. In brief, he argued that health care policy increasingly guides mental health care policy. A key flaw in health care policy is the predominance of general hospitals, which creates an emphasis on acute care and inpatient care. To the extent that mental health becomes similarly consumed by the hospital, mental health policy dooms itself to the same failures as health policy. By demonstrating a key way in which mental health and health care policy are intertwined, this insightful article pointed to the challenges facing mental health policy and American psychology as health care undergoes reform at the national level.
- Over the past decade, and particularly in the past several years, health services researchers, health care providers, policy analysts, and policymakers have been discussing health care reform. Several prominent proposals exist (e.g., Enthoven & Kronick, 1989a, 1989b; Himmelstein et al., 1989), and President Clinton's clear commitment to health care reform makes a serious attempt at some type of major and systemic change inevitable. It was only recently, however, that a proposal emerged for the construction of a private-sector mental health benefits package designed to serve as a model in the public debate over health care reform (Frank, Goldman, & McGuire, 1992).
- Any significant change in health care policy will pose serious challenges and opportunities for mental health care policy. Reform may bring the opportunity to remedy flaws in both health and mental health policy. Reform may also create new flaws that could lead to failures in mental health policy. Psychologists will face decisions about the kinds of roles they will take in the debate over both health and mental health care reform. To illustrate these challenges, opportunities, and decisions, I will provide one example of a proposal for health care reform and one example of a proposal for mental health benefits designed to fit within national health care reform.
- Of the existing proposals for health care reform, the one currently receiving the most attention is a plan of "managed competition," championed over the past decade by Enthoven and Kronick (1989a, 1989b). The fundamental tenets of managed competition can be given only the most cursory treatment here; interested psychologists—and most psychologists ought to be interested—should read Enthoven and Kronick's detailed articles. In this plan, all persons not covered by an existing public program (e.g., Medicaid and Medicare) would be able to buy subsidized coverage through either their employers or "public sponsors" if they are not employed, are self-employed, or are employed in a small business. Public sponsors would be created at the state level through federal fiscal incentives. These sponsors would serve as brokers, selecting coverage and contracting with health plans. In this way, the buying power of small employers and individuals would be enhanced to resemble those of very large employers. Medium and large employers would be required to cover all full-time employees by contributing at least 80% of the cost of coverage. Several taxes, designed to encourage employer participation and to raise revenue, are a part of the plan.
- In sum, this plan mixes public and private financing. It allows for competition among large health care providers and plans but gives substantial fiscal resources to the federal and state governments that purchase these plans. As a result, government will have leverage to develop a regulatory framework that would require the plans to operate in the public interest. This latter point is critical to mental health policy. The competitive aspects of managed competition would encourage providers, such as HMOs or other organizations, to create efficient, low cost operations in order to be more attractive to the large employers or public sponsors who purchase the plans. One way to reduce costs is to reduce benefits for certain kinds of care. Mental health care is certainly one likely target of cost cutting and, in fact, is not provided for in more than one third of the current legislative proposals for health care reform.
- The federal government, under managed competition or other versions of health care reform, could mandate or encourage mental health benefits. Already, the fight is being joined. The mental health section of the American Public Health Association, for example, is lobbying for parity between health and mental health benefits. The recent article by Frank et al. (1992) outlined how such mental health benefits might actually be designed by modeling a rational package of benefits intended to assure access to treatment for persons suffering from mental disorder. Although including mental health benefits in a national health care package makes obvious sense, there are still risks involved. Kiesler's (1992) article sounded a cautionary note: Incorporating mental health benefits within a national health plan may result in a mental health policy excessively based on medical models.
- If mental health benefits become part of a national plan for health care, bloody fights will ensue over the shape and breadth of such benefits. As one example, the Frank et al. (1992) plan provides unlimited ben-

Doomed and Double Doomed? Health Care Reform and the Mental Health Policy Challenge

Abram Rosenblatt
Department of Psychiatry,
University of California, San Francisco