Academy Panel Joins the Fray Over Job Testing

It endorses the controversial practice of reinterpreting scores of blacks and Hispanics on a widely used employment test

A PANEL of the National Academy of Sciences has endorsed a system for reinterpreting the ability test scores of blacks and Hispanics to make them competitive with those of whites.

This controversial recommendation is part of a report attempting to resolve a dispute between the Labor Department and the Justice Department over the use of the General Aptitude Test Battery (GATB), the most widely used civilian employment test in the country.

Job candidates who take the GATB at state-run employment services are referred to employers according to a "race norming" formula that helps employers identify the highest scorers within ethnic categories. The practice, promoted by the Labor Department, has been attacked by the Justice Department as "intentional racial discrimination." However, the academy panel, headed by Yale University statistician John Hartigan, concluded that the practice is justified because of the imprecision of the test.

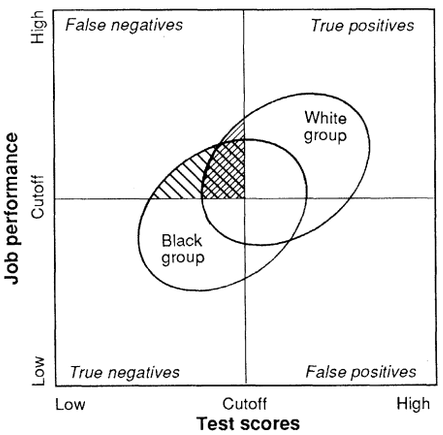
The report,* issued on 22 May, has a direct bearing on two different but often intertwined issues: the value of ability tests in predicting a candidate's future performance on the job, and appropriate strategies for minority applicants, who argue that tests unfairly discriminate against them.

While research has shown that "objective" selection procedures such as the GATB are the best single predictor of employee performance, it has also shown that such tests put blacks and Hispanics at a severe disadvantage. The resolution to this conflict between minority interests and concerns about worker quality will affect millions of job referrals at state employment services and will have implications for how tests are used throughout private industry.

In coming years, the problem is likely to escalate: the country's rapidly changing demographics means that the majority of future work force entrants will be minorities. Meanwhile, test use is on the rise. Since the debate cannot be resolved scientifically, no matter what the government decides, the ultimate outcome will depend on the courts.

The focus of this particular controversy, the GATB, tests three major domains: general cognitive ability, perceptual ability, and psychomotor ability. Since 1981, the Labor Department's Employment Service has been encouraging its use to improve screening at state employment services, which annually process about 20 million job-seekers and refer them to private employers.

To offset the fact that blacks and Hispanics have lower scores than whites on the GATB (as they do on virtually all standardized ability tests), the government has promoted within-group scoring. Applicants label themselves as "black," "Hispanic," or "other," and their test scores are recomputed to reflect their percentile ranking within their racial group. Top scorers from all groups are then referred to employers.



How good a predictor? *Many potentially able workers fail the GATB test (false negatives), and since blacks as a group score lower, they are more likely to be in this category.*

In November 1986, the Justice Department ordered the employment service to "cease and desist" from the practice. The service agreed to put a moratorium on expansion of the program pending the completion of the study. (The stay is continuing as the government studies the findings.)

The Academy's Committee on the GATB was charged by the Labor Department with assessing whether the test is a good one, whether it is valid (a useful predictor of job performance), whether it is fair to minorities, and whether race norming is a good idea. The answer on all counts was yes.

By supporting GATB, the academy panel was endorsing a concept that was formulated in the late 1970s. That was when psychologists Frank Schmidt of the University of Iowa and John Hunter of Michigan State University came up with an answer to the landmark 1971 decision of the Supreme Court that had almost crushed testing. The case, *Griggs* v. *Duke Power Co.* established that any employment practice having "adverse impact" on minorities constituted evidence of discrimination. This put the burden of proof on employers who had to show that their criteria were directly job-related. Many employers abandoned ability testing rather than devote the enormous resources necessary for constructing and validating job-related tests.

Until the late 1970s, psychometricians could see no way around the need for separately validated tests because the results of validation studies were so varied that it did not appear that those from any one study could be generalized from one place to another.

But then "validity generalization" emerged on the scene. This concept means, simply put, that a general measure of cognitive ability that is valid for some jobs is valid for all jobs. The theory is based primarily on the work of Schmidt (formerly at the U.S. Office of Personnel Management) and Hunter, who applied new analytical techniques to 500 validity studies. They found that when the results of the studies were corrected for various distortions—primarily those imposed by small study samples—they yielded substantial correlations with a wide range of jobs. Their conclusion: "professionally developed cognitive ability tests are valid predictors of performance on the job and in training for all jobs . . . in all settings."

The employment service was excited by these findings, particularly in light of Hunter and Schmidt's calculations that widespread adoption of the GATB would result in an $80-billion-a-year savings to the economy through increased productivity. The services implemented a pilot program to test out the concept.

Because blacks and Hispanics get lower scores, they added within-group scoring to achieve parity in referrals. For example, because blacks as a group score a standard deviation below whites, the raw scores of those who fall in the 50th percentile are assigned to the 84th percentile. Hispanics with the same score are in the 66th percentile. As of 1986, when the Justice Department stepped in, the system had met with

*"Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery" (National Academy Press, Washington, D.C., 1989).

considerable favor from employers and was being used in 40 states on about 8% of applicants.

The NAS panel members generally affirmed the whole program. They confirmed that the GATB is a valid predictor of job performance, although they concluded that Hunter and Schmidt's estimated correlation of 0.5 between test scores and job performance was too high. Citing additional, more recent, studies, they adjusted it down to the neighborhood of 0.3. Although many critics of ability testing believe a 0.3 correlation represents an improvement over random choice so small as to be meaningless, panel leader Hartigan said at a Labor Department briefing that the critics have misinterpreted the effectiveness data, and that use of the tests actually improves the ability to predict a worker's productivity by 30%. And he reaffirmed the panel's agreement with Hunter and Schmidt that the GATB is more reliable than any other single selection criterion, including interviews, educational background, skills, and job experience. Said Hartigan, "we probably cannot afford not to use" the GATB.

The most crucial aspects of the report have to do with questions of whether blacks and Hispanics are unfairly dealt with, and what to do about the fact that their scores are significantly lower than those of whites. With regard to whether the GATB is racially biased, the NAS endorsed findings from a vast body of research on the subject showing that it is not—that is, the tests predict equally well for blacks and whites. The NAS confirmed that, if anything, the test slightly favors blacks by "overpredicting" their job performance.

Many test critics, including Richard T. Seymour of the Lawyers Committee for Civil Rights, have claimed that the test is racially biased because more potentially able black workers are rejected by the test, and more poor white workers get passing scores. In scientific terms, more blacks than whites fall in the category of "false negative," and more whites are "false positives."

The NAS panel said, however, that this disproportion has nothing to do with race per se but arises from the fact that it is the marginal scorers who are most likely to fall in the false negative category (see graph). This can be demonstrated by performing the same analysis using one racial group.

As the academy panel pointed out, the problem of false negatives is an inevitable result of the limited predictive capability of the test. But the panel has put itself in a somewhat awkward position. Study director Alexandra Wigdor emphasized that "this correction is not for racial underprediction, it is underprediction for low scorers." But the report presents race norming as a way to "ensure that able black and white workers have the same chances of referral"—thus implying that the test is biased.

What the academy has done is to take a remedy adopted by the employment service on purely pragmatic grounds and present it as one that is scientifically justified—even though, according to James Sharf, an industrial psychologist at the Office of Personnel Management, the vast bulk of research shows that pure rank-ordering of scores "is the only scientifically justified position." Sharf, a member of the committee's liaison group, quotes Hartigan as saying, at a meeting 2 years ago, that "this committee is not about to put a scientific fig leaf on a naked political argument." Sharf says the feeling at OPM is that the committee has done just that. Hartigan could not be reached for comment.

The widespread adoption of race norming could open up a Pandora's Box of new questions and litigation. Nonetheless, overt and systematic policies of racial preferment may be better than informal arrangements that are often neither efficient nor fair. As Wigdor observed, employers are in a "tremendous bind" because they risk adverse impact suits when they use objective selection procedures, and reverse discrimination suits when they set up programs favoring minorities. As a result, "a lot have turned to quiet, unobtrusive quota systems that can't be recognized in court."

■ CONSTANCE HOLDEN

Consorting on Superconductors

They may be the most powerful corporate rivals in U.S. research, but IBM and AT&T have decided to join forces—along with the Massachusetts Institute of Technology and Lincoln Labs—to ensure American primacy in superconductivity in the 21st century. The venture, to be known as the Consortium for Superconducting Electronics, will attempt to transform what has been largely an interesting laboratory phenomenon into real-world applications. If it appears to be working, it could become something of a model for corporate rivals in other fields to work together with universities on long-term applied research programs.

The initial focus of the consortium will be applications in the world of microelectronics, such as high-speed signals processing circuits and junctions between electronic devices, that are expected to constitute the first uses of the new superconductors.

This may prove particularly wise because superconducting electronic devices are expected to be less affected than many other putative applications by the recently reported (*Science*, 26 May, p. 914) phenomenon known as "flux creep" that can destroy the superconducting properties of the new materials when they are exposed to magnetic fields. Still, the most promising electronics applications are, as yet, uncertain. Says William Brinkman, director of physics research at AT&T's Bell Laboratories, the consortium should "find an answer to the question of whether there are technical opportunities open to us."

Indeed, the fact that the big players in high-temperature superconductivity have decided to join forces is being viewed by some as an indication that they are looking for a way to share some of the costs while they explore the formidable barriers that lie before them. Says Dean Eastman, a vice president of IBM's research division: "We believe that it's going to take considerable time to achieve applications, so we need to look at this over the long haul, not just when it's in vogue among scientists."

A novel feature of the consortium that sets it apart from other university-industry research arrangements is that it is built around a detailed plan, complete with technical milestones, and it will be managed by a central group to be located at MIT. "It is not a consortium in which IBM, AT&T, Lincoln Labs, and MIT are each following their own programs and sharing results; they will be following a single technical plan," says MIT provost John Deutch. Adds Eastman of IBM, "the consortium will act like a small company."

Not so small, though, when ranked against other superconductivity start-ups. Indeed, the new entity will command an annual budget of $12 million to $15 million a year. A grant of $4 million to $6 million is being sought from the Defense Advanced Research Projects Agency to finance work at MIT; the rest will be kicked in by each industrial partner. Each institution will have the equivalent of five or six full-time researchers working for the consortium.

Deutch says he will be spending some time over the next year seeking additional members for the consortium from industry, the national laboratories, and other universities. Similar consortia could follow, Deutch predicts. "We have it in mind as being a model for how universities, industry, and the national labs can work together on things that are in the national interest."

■ COLIN NORMAN