DOCUMENT RESUME

ED 284 894 TM 870 441

AUTHOR	Baker, Eva L.
TITLE	Critical Validity Issues in the Methodology of Higher Education Assessment.
PUB DATE	86
NOTE	9p.; In: Assessing the Outcomes of Higher Education; see TM 870 436.
PUB TYPE	Speeches/Conference Papers (150) -- Viewpoints (120) -- Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	College Students; *Educational Assessment; Educational Testing; Evaluation Criteria; *Evaluation Methods; Evaluation Problems; *Evaluation Utilization; Higher Education; *Measurement Objectives; Outcomes of Education; *Student Evaluation; *Validity

ABSTRACT

Ways to improve the validity of assessment of college students are discussed. Validity problems often have occurred because the purpose of assessment was not clear. Traditionally, college students have been tested for admission and placement. However, the results of these tests have been used for other purposes, such as comparing institutions. Whatever the intended purpose of testing, the findings from assessment are often used for something else. College students are commonly assessed for accountability, certification, or institutional self study. Assessment is an indicator of quality and is also used as an intervention for educational change. Individual instructional assessment (IIA) is proposed as an assessment model which incorporates assessment with the institution's teaching mission. IIA uses assessment to recognize and extend individual student accomplishment. To implement the IIA system, commitment from administrators and faculty would be necessary. In addition, the measurement community would need to provide assistance, including tests, technological support, training, and recommended procedures. (GDC)

Critical Validity Issues in the Methodology of Higher Education Assessment

Eva L. Baker
UCLA Center for Student Testing, Evaluation, and Standards

Paper presented at the forty-seventh ETS Invitational Conference, sponsored by Educational Testing Service, at The Plaza, New York City, on October 25, 1986.

Critical Validity Issues in the Methodology of Higher Education Assessment

Eva L. Baker
UCLA Center for Student Testing, Evaluation, and Standards

Validity is the grand old concept of assessment. It stands for a complex set of ideas involving the purposes of assessment, the match of information obtained to such purposes, and the process by which information is verified. Validity in testing, as in English, is about truth. This paper focuses on increasing the validity of student assessment in higher education.

Since validity is an apparent good, why do we have a problem with it in higher education—or anywhere? Our validity problems occur because we frequently are unclear about the purpose we are serving with our assessments, a situation that also clouds the inferences we should make from our findings.

Traditionally, at the postsecondary level, we have tested students for admission and placement. Admissions testing has drawn public attention because of its centrality in the allocation of equal educational opportunity and because the average admission test score has become a shorthand description for the educational standards of colleges and universities—the purported goodness of the education directly related to the difficulty of admission. More recently, average admissions test score has been applied, in a similar way, to evaluate the precollegiate educational effort. Although it has been common at private schools to judge educational quality in terms of the number of students admitted to the most elite postsecondary institutions, it was only relatively recently that such college admission test scores were used to compare state educational systems (U.S. Department of Education, 1985). Both uses of admissions tests raise obvious problems relating to the validity of inferences: are we talking about the quality of the educational institutions themselves, the quality of their clients, or some unknown combinations of the two? Furthermore, such quantitative shorthand whets some appetites for other

simplified measures of educational quality. So, of increasing interest to the postsecondary community and those compelled to comment about its effectiveness, is the utility of student achievement measures for assessing postsecondary educational quality. Driving these interests in student assessment are legitimate public concerns about higher education costs and benefits. The spate of attention to this issue by the Federal establishment was perfectly predictable: as precollegiate educational programs were shifted to States for management, the majority of the remaining federal educational investment was directed to postsecondary students. Accountability went to college.

Present Methods

From all reports, each of the existing systematic assessments of student academic performance in colleges and universities has developed through top-down mandate. How high up that top is varied, with the present ceiling at the statehouse. The intended purposes served by such mandated student assessment include accountability (reporting to legislatures), certification (verifying performance for existing teachers), or institutional self-study (McClain and Krueger, 1985). Although assessment systems may begin with one ostensible purpose (who goes to what segment of higher education), a mutation such as outcome assessment is not hard to imagine. A major fact about testing is that whatever its original purpose, the findings from assessment are always used for something else.

From all appearances, many existing assessments of postsecondary students share the methodology and flavor of precollegiate, large-scale testing activities. The measures are standardized. They are formulated for and administered to the group. They often focus on minimums. They have great symbolic value, and their functional value is unknown. To the extent that student assessment measures become widespread, I will predict that their original purposes will be transformed and that they will also drive out other indicators used to evaluate comprehensively the quality of higher education institutions. Simply look at precollegiate education as relevant history. Mandated, large-scale testing occurred because the precollegiate system had no convincing information about its quality. No information was available to refute claims that kids couldn't read and write, let alone do fractions and analyze Shakespeare.

Assessment as an In[illegible]nal Reform

At the heart of this disc[illegible]ent as a bureaucratic tool. Bureaucracies seen[illegible]nt program serving at least two purposes: firs[illegible]em quality; second, and increasingly more impor[illegible]ion. In precollegiate education, for instance, imp[illegible]testing is seen in itself as a major educational refo[illegible]way to measure the effects of changes in educational s[illegible]a classic quick fix. The rhetorical benefits of formal asse[illegible]ticulate standards, focus instruction, motivate students a[illegible]eet to the fire, etc. The feared costs of such assessments incl[illegible]g to trivia the important goals of education, increasing the dropout rate, generating systematic attempts to "get around" the mandate, narrowing the curriculum, and so on. Studies of acutal effects of testing reforms will be released shortly and some light may be shed on the utility of assessment as a productive instrument of educational change.

Assessment as a Quality Indicator

The use of student achievement is a legitimate important indicator of educational quality. If they are to be used as part of a system of higher education, student assessment programs must be constantly held to their purpose: to provide an accurate and representative reflection of educational quality. Methodology used in student assessment does not meet this purpose. In my view, student assessment programs must intrinsically relate to real instructional programs in departments and courses. They must reflect the diversity of our offerings and what students learn from their coursework and their college experience. At present, we have relatively little evidence to document the effects of our educational efforts in higher education. I believe we can collect such evidence in a way that will avoid the bureaucratic and irrelevant character of much top-down assessment. We should try to avoid the use of omnibus assessment, where a single instrument is purported to be a major valid indicator of quality. The nature of higher education is such that using a single common measure to reflect student learning will provide very little valid information about educational quality. Most everything will be missed. We may, better still, find a way to use student performance assessment as a powerful instrument of improvement.

Developing an Approach to Individual Instructional Assessment

The model for student assessment in higher education I propose is one that incorporates student assessment as part of the teaching mission of the institution (Cross, 1986). Its purpose is to contribute to the development of educational quality. Call it individual instructional assessment (IIA). IIA develops from a view that colleges and universities have teaching responsibilities to individual students. The teaching responsibilities for individual students get executed as students relate to one another, to professors, to teaching assistants, and to other institutional resources. The product of this individual experience is what we should assess. Even though teaching is sometimes a mass act, its reality occurs in the complex interaction among the students and all these resources (Pace, 1985). To acknowledge and assess the individual, distinct, personalized nature of this experience is critical. However, such acknowledgement should not be confused with models of instruction (such as those advanced and tested by Keller (1969) and Bloom (1967; 1984). IIA does not presuppose self-paced instruction and is independent of instructional strategy. The purpose of IIA is to use assessment as a way to recognize and extend individual student accomplishment rather than to homogenize it. Its slogan was promulgated by Judah Schwartz (1978), in other contexts, some years ago: "People cor in groups of one." So do higher education institutions.

A new approach to student assessment in postsecondary education is needed. This approach would use as its centerpiece the specific accomplishments of students in academic courses and courses of study, instead of their performance on specially constructed, mandated measures. So I will not discuss today a procedure to develop particular instruments. Outcomes of higher education would be documented by providing a wide range of *examples* of the kind of work accomplished by students at various levels and majors. The system would not be uniformly applied to all courses, nor would exhaustive reporting be expected. Rather an *institutional portfolio* would be created. If numbers are required, as they almost always are, frequencies of students performing at the illustrated level or above would be provided for the academic majors assessed. It is bottom-up demonstration of quality, clearly superior, I think, to judgments made on the basis of transcript analyses or catalog review.

The characteristics desired of such measures are obvious. The common, casually developed tests of knowledge and information in rampant use could realistically provide only a piece of the information. New, carefully

developed tasks for essay examination or term papers would be prepared. Criteria for judging the quality of responses would also be articulated. In operation, these assessments would be administered on a schedule naturally demanded by course organization. Feedback to students would be provided rapidly and in a way that strengthens the personal nature of the college experience.

What those tasks should be and the form of the feedback should be a faculty matter. Educational quality, in terms of what and how well students learn the full range of academic offerings will thus be directly affected. As present institutionally-generated student assessment is focused on scheduled, quantitative summaries of students' performance, IIA is periodic, qualitative, formative, diagnostic, and informative. IIA would also serve to increase rather than decrease the range of approaches used to assess learning. It also has particular strength as a means to provide careful differentiated feedback for students.

Of course, such a position requires a massive effort to train faculty members. They need to see that the way they assess students communicates what they view as important to learn. They need to believe that careful, timely, and personalized feedback can transform the college experience for students. They need to see assessment as more than a means to grade students or to meet bureaucratic requirements. It must contribute to their teaching effectiveness.

Do faculty care enough to engage in the serious work of developing high quality measures of course performance? We know they are relatively unskilled now. Whether some would embrace the use of high quality measurement approaches (such as domain-referenced assessment) remains to be seen.

What conditions are required for such a system to work?

- Agreement from top management that such an approach would directly rather than indirectly both impact and reflect higher education quality and that it is worth doing and superior to approaches using single measures.
- Incentives for faculty to take this responsibility seriously.
- A plan for institutional development, first to find leading academic institutions willing to undertake a pilot effort, and, within institutions, prestigious academic departments to provide the model for others.
- Useful approaches, tools, and training procedures from the measurement community.

Necessary Contributions from the Measurement Community

Colleges and universities, if they were to take seriously and systematically the charge to improve educational quality, need certain assistance from the measurement community. For example, approaches to the measurement of deep understanding of subject matter would need expansion. In a project in this domain we are attempting to develop procedures for assessing essays and term papers that incorporate appropriate cognitive representation of subject matter (Baker & Herman, 1986), reliable, and valid scoring of student responses, and procedures that do not demand inordinate time to evaluate each student's effort (Quellmalz, 1984). The measurement community needs to expand the options it offers college professors to assess subject matter and cognitive understanding.

Secondly, technological supports to the development of assessments are at least on the drawing board (Baker & Linn, 1985). The search should intensify for procedures to use computer technology to represent subject matter knowledge and to develop locally appropriate measures of student performance. As part of new OERI Centre for Research on Testing, we have a design project to explore techniques from artificial intelligence to create a test developer assistant (Baker, 1986).

Third, help from offices of institutional research and evaluation is needed to provide the structure and training required for such an experiment to work.

Summary of Potential Effects

If successful, the results of IIA should be:

- to deepen the sense of intellectual engagement of students by requiring of them high level, defensible performance, and by providing timely individualized feedback,
- to stimulate faculty reflection on the real teaching mission of colleges and universities,
- to avoid the use of marginally valid measures in the assessment of higher education, and
- to provide appropriate indicators of higher education quality, in the form of institutional portfolios.

In this way, we can contribute to the responsible assessment of our higher education institutions. We must recognize that our institutions are complex, our students are different, and that our assessment approaches need to reflect those complexities.

References

Baker, Eva L. "The Impact of Advances in Artificial Intelligence on Test Development." *Continuation Proposal for OERI Center for Student Testing, Evaluation and Standards*, Los Angeles, CA: UCLA, 1986, pp. 82-87.

Baker, Eva L. & Joan Herman. "Issues in Content Assessment." *Continuation Proposal for OERI Center for Student Testing, Evaluation and Standards*, Los Angeles, CA: UCLA, 1986, pp. 74-81.

Baker, Eva L., & Robert L. Linn. *New Testing Technologies*. Los Angeles, CA: Office of Technology Assessment Study on Standardized Testing, UCLA, 1986.

Bloom, Benjamin S. "The Search for Methods of Group Instruction as Effective as One-on-One Tutoring." *Educational Leadership*, May 1984.

Gross, Patricia K. "A Proposal to Improve Teaching." *American Association for Higher Education Bulletin*, Vol. 39, no. 1, Sept. 1986.

Keller, F.S. "Goodbye Teacher." *Journal of Applied Behavior Analysis*, 1968, 1, 79-89.

Pace, C. Robert. "Perspectives and Problems in Student Outcomes Research." *Assessing Educational Outcomes*, 47, 1985.

Quellmalz, E. "Designing Writing Assessments: Balancing Fairness, Utility and Cost." *Educational Evaluation and Policy Analysis*, 6(1), 1984.

Schwartz, Judah L. "Assessment that Respects Complexity in Individuals and Programs." Paper prepared for the National Conference on Urban Education, held in St. Louis, Missouri, July 10-14, 1978.

U.S. Department of Education. *Indicators of Education Status and Funds*, Washington, DC: P. A-6, 1985.