DOCUMENT RESUME

ED 285 878 TM 870 258

AUTHOR Hadley, Marilyn; Vitale, Patrick
TITLE Evaluating Student Achievement.
PUB DATE 85
NOTE 14p.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Academic Standards; College Entrance Examinations; Evaluation Methods; *Evaluation Problems; *Grade Inflation; *Grading; Graduation Requirements; Higher Education; High Schools; Measurement Objectives; Standardized Tests; *Student Evaluation; Student Promotion
IDENTIFIERS Scholastic Aptitude Test; Test Score Decline

ABSTRACT

The validity of grades in higher education as a measure of what a student has actually learned has been a concern to both the public and academicians for over three decades. This was one of several issues discussed in a report by the National Institute of Education Study Group on Conditions of Excellence in American Higher Education. Furthermore, the problem has been complicated by grade inflation since the 1960s. As a result of this vagueness in the meaning of college grades, states have become more involved in college student assessment, especially through the use of standardized tests. It is recommended that: (1) colleges and universities continue to monitor grade inflation; (2) colleges consider changing from a five-point to a 13-point grading scale; (3) colleges consider the use of criterion-referenced grading rather than norm-referenced grading; (4) state agencies involve university and college faculty in studying and adopting changes; (5) standardized tests be developed at the state or local level, since the United States does not have a national curriculum; and (6) an instructional model be used to assess students before college entry, during the undergraduate program, and at graduation. (JGL)

EVALUATING STUDENT ACHIEVEMENT

Marilyn Hadley

Patrick Vitale

Marilyn Hadley and
Patrick Vitale

EVALUATING STUDENT ACHIEVEMNT

The focus of the public examination of education has shifted in the past year to higher education with the release of the report by the NIE Study Group on Conditions of Excellence in American Higher Education (NIE, 1984). The purpose of this group was to make suggestions for improving higher education, particularly at the undergraduate level. The report identifies, analyzes and makes recommendations for meeting three conditions of excellence: student involvement, high expectations, and assessment and feedback.

It is not surprising that the area of assessment is identified as a major concern by the NIE Study Group. Historically, assessment of student achievement has been a concern of both the public and academicians. In the early 1950's, a movement began to improve college and university teaching. A study of teaching practices at the time indicated that faculty skills and efforts to construct reliable tests were limited, thus raising concerns about the use of grades (Umstattd, 1954). Umstattd voiced a concern about measuring learning versus giving grades, and recommended careful study of this problem by institutions.

Becker, Geer and Hughes (1968) echoed the concern of the conflict between learning and grades in their study of college students. As result of a two-year ethnographic study of undergraduates at the Unive ty of Kansas, they found that grades were the major institu 1 "valuable." Personal intellectual growth and scholarship were impo to some students but were not viewed as universally valuable as grades. They also found that students were able to get good grades without necessarily learning. In the students' opinion, success was measured by a good grade point average. Becker et al. (1968) concluded that the anti-intellectualism of grading results in a dilemma of how to reward true achievement rather than grade-getting skills.

According to Oldenquist (1983), attitudes about testing as evil and grading as a way of labeling individuals as successes or failures were a result of the social turmoil of the 1960's and 1970's, an era when the rights of women, blacks, disadvantaged, and handicapped were promoted.

It is during this time that scores on the Scholastic Aptitude Test began an 18-year decline, and that grade inflation at all educational levels began. Oldenquist attributes this "decline in education" to a reluctance of educators to apply strict standards because they could not see the difference between the elitism implied by standards and the elitism of social class or privileged group.

That grade inflation occurred from 1964 to 1974 is well documented (Bejar & Bleu, 1981; Oldenquist, 1983). Grade inflation, which is generally viewed as a progressive rise in grade point average without a concurrent rise in student ability as measured by college entrance exams (Bejar & Bleu, 1981), contributed to the public lack of confidence in the traditional evaluation system. The reliability of grades at the undergraduate level was found to be affected by grade inflation when a five-point scale (A, B, etc.) was used (Millman et al., 1983; Singleton & Smith, 1978). The reliability was not affected when a thirteen-point scale (plus and minus grades) was used.

Current Concerns

In view of the previously cited literature, it is apparent that dissatisfaction with grading and evaluating students has been an issue for more than three decades. What is new in the 1980's is a change in the focus of concern. Today, standardized tests are being used increasingly as a way of evaluating students; the level of intervention has shifted from the local institution to state officials and groups, and the focus of standardized testing is shifting from measuring entry level ability to measuring learning of students who graduate.

Current Practices

In order to meet the condition of excellence for assessment feedback, the Study Group (NIE) calls for the establishment and maintenance of high standards of institutional and student performance. It calls for entry standards to be identified and publicly stated in terms of student knowledge and skills, as opposed to the use of cutoff scores on standardized tests of high school grade point averages. It also calls for a measurement on student outcomes in terms of knowledge and skills as opposed to the accumulation of a given number of credits because credits are "measures of time and performance, but they do not indicate the academic worth of course content" (p. 13). The purpose of this paper is to describe current evaluation practices used in undergraduate education at entry, during a student's program, and at graduation. The authors will then recommend what can be done to assist institutions in responding to the concerns about evaluation.

Entry Level

Admission criteria for undergraduate programs generally include some combination of college entrance examination scores, high school grade point average or rank in class, and specific type and number of

high school courses. A review of admission requirements for sixteen public and four private institutions in five Upper Great Plains states was conducted by the authors using The College Handbook 1984-85 (The College Board, 1984). Most institutions require high school class rank or standardized test scores (ACT is preferred, SAT is accepted). Three private institutions consider both class rank and test score along with the high school record. Four of the institutions, all public, still maintain open admission for resident students.

As mentioned previously, the use of college entrance exam scores is generally accepted as one criterion for admission. The mean SAT and ACT scores have declined from the mid 1960's to the present. However, this trend seems to have been reversed .

Institutions have also used standardized tests for counseling and placement purposes. These efforts have been strengthened by new initiatives taken by legislatures during the past few years (Mingle, 1985). For example, the Florida legislature has mandated that entry tests of basic computation and communication skills be used as screening devices, and students who require remediation enroll in community college "college prep" courses. In New Jersey, all entering freshmen and transfers take a basic skills test of reading comprehension, sentence sense, computation and elementary algebra. Test data are used for course placement and counseling. In Ohio, high school juniors are tested on writing, science readiness, and math skills. Students receive feedback from the college of their choice in time to take corrective action in their senior year.

In summary, requirements and standards used for admission generally include high school grade point average or rank in class, successful completion of certain high school courses, and/or college entrance exam scores. The trend has recently been to increase the requirements and standards. Tests are also increasingly used to counsel and place students in courses.

Evaluation of Student Progression

Once a student is admitted into an undergraduate program, s/he is generally evaluated by professors as a part of an individual course as well as in terms of progression in an academic program. Within courses, professors collect information about student behavior in order to assign grades. In addition, professional programs generally set standards for admission into their upper level program, which is usually the last two years of undergraduate work.

The formal evaluation of the student is generally recorded using a five-point scale, ranging from 4=A to 0=F (Millman et al., 1983). A student's performance can be measured in a variety of ways, including tests, papers, classroom discussion, lab work and attendance. The emphasis on tests and classroom contribution in the 1950's (Umsttatd, 1954) has changed somewhat over time to include an emphasis on papers (Barnes, 1984). Assessment is continuous, often with weekly assignments and several tests prior to the final exam.

Becker et al. (1968) describe the teaching/learning situation "as actually an exchange of reward for performance, rather than as some kind of an educational process" (p. 63). They propose that a classroom contract exists on an informal basis, the terms of which are discovered by the student via the course syllabus, the professor's actions and words, and information from students who have had the course previously. Expanding on this idea of a contract, Barnes (1984) suggests that students actively engage faculty in the process of grading and that the final grade is a consequence of negotiations between the student and professor.

In an Educational Testing Services study of grades during the 1960's and 1970's, it was found that colleges and universities initiated changes in their grading policies to include pass/fail and credit/no credit options (Suslow, 1976). Incompletes were increasingly given in place of D's and F's, and grading systems were adopted that included the assignment of greater than four points for A+ (e.g., 4.3 or 4.5).

During the period 1964 to 1974, Suslow (1976) found that the percent of A grades doubled from 16% to 34% while the percent of the C grades fell from 37% to 21%. The mean grade point averages during those same years consistently increased from a grade point average of 2.4 to a grade point average of 2.95, a rise of .55 points (Figure 1). During this same period when the GPA's increased, the SAT's decreased (Figure 2). Oldenquist (1983) reported that the mean SAT score in mathematics declined from 502 in 1963 to 466 in 1981, while the verbal score showed a greater decline during the same period of 480 to about 420 (Figure 2). Grade inflation is described as a progressive rise in undergraduate grade point averages accompanied by a continuous decrease in SAT and ACT scores which are measures of student ability (Bejar & Bleu, 1981).

Since 1974, the grade point averages have tended to level out and seem to have remained fairly consistent, indicating that grade inflation has been arrested. However, it does not mean that grade inflation has ended. In contrast to 1964, we are still in a period of grade inflation.

A number of reasons have been suggested for the dramatic rise in the undergraduate grade point average. Suslow (1976) speculated that these include changes in the type of students, faculty behavior due to that change, and innovations in the grading system. Other factors that may have contributed to the increase in grade inflation include increased student participation in academic policy formation, selection of the credit/no credit option to avoid grades in difficult courses, contract learning, and student evaluation of faculty performance. A concern on the part of college and university administrators for decreasing student enrollments, a general relaxation of academic rules, allowing all students to enter, and a reduction in the depth and breadth of course requirements in the freshman and sophomore years could also have contributed to an increase in grade inflation.

Researchers (Komorita & Graham, 1965; Masters, 1974) have found that the reliability of grades is directly related to the number of response categories used. Since grade inflation results in a decreased

Figure 1. Comparison of a number of university mean GPA's with USD for entering freshmen

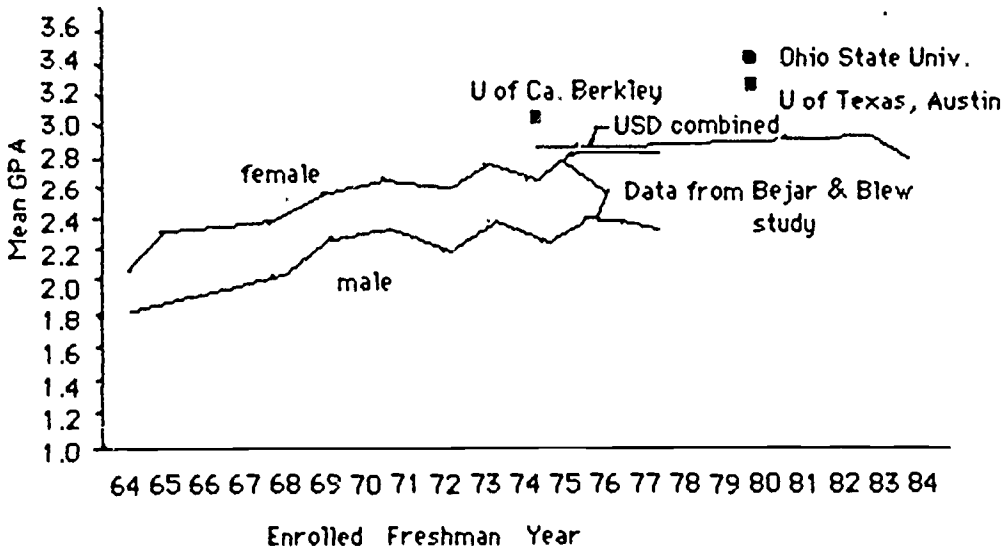
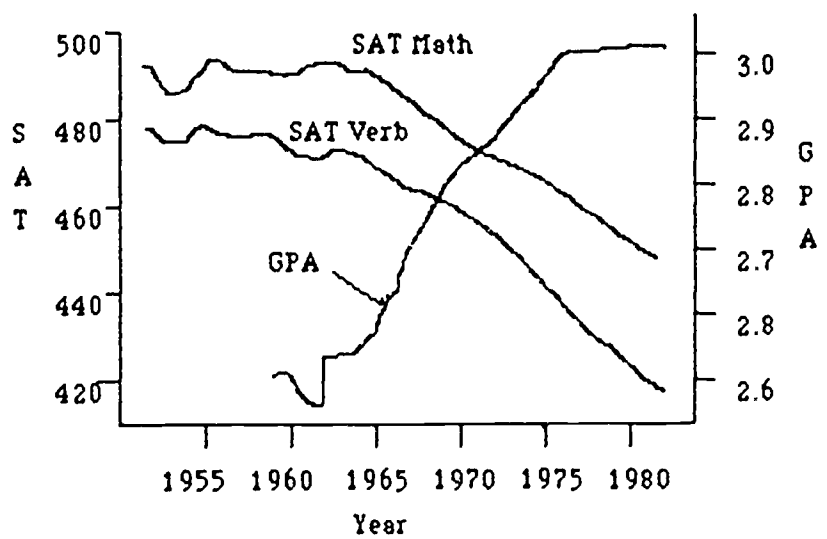


Figure 2. Comparison of mean SAT Verbal and math scores with GPA for the 1954 to 1981 period



number of categories used, reliability decreases (Figure 3). Average intercorrelations of students' grades have declined as grade inflation progressed, as a result of using a restricted number of grade categories. That the trend reversed itself in 1971 could have been caused by the introduction of plus and minus grades as shown in Figure 3. Grade inflation during this period of reversal of the average intercorrelation continued, while the reliability of grades rose to higher levels than the period before grade inflation.

Oldenquist (1983) reported that Colleges of Education were the leaders in grade inflation, followed closely by the Schools of Social Work. He reported that the average undergraduate GPA given by the University of Texas at Austin's Education College was 3.3, compared with a 2.4 in business, and that in the spring of 1977, 45% of the grades given in undergraduate courses by the College of Education were A's compared with 15% in business and 31% in the humanities. He reported a similar trend at Ohio State University (Figure 1).

Institutions are aware that grade inflation has occurred, and are beginning to monitor grade distributions (Mingle, 1985). Hambleton and Murray (1977) surveyed faculty and student views concerning the uses of grades in different instructional settings and the appropriateness of grading systems in common use for accomplishing the intended uses of grades. The major conclusion of the faculty and students was that a criterion referenced grading system was more desirable for evaluating course outcome than a norm referenced grading system.

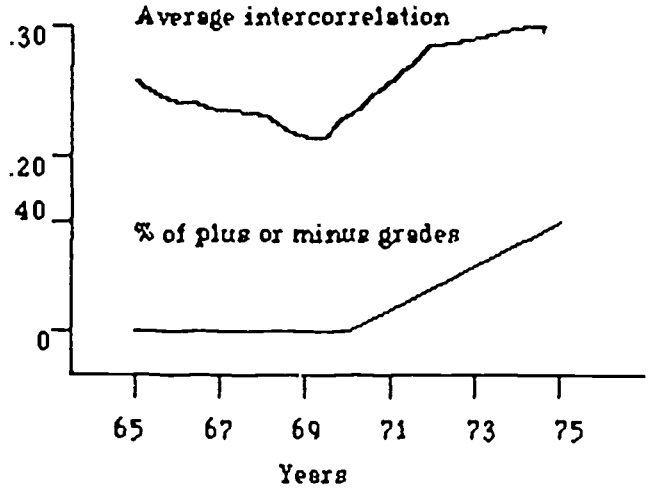
Despite the concern about grade inflation, grades continue to be the traditional measure of student achievement in courses and student progression from general course work to professional schools. A recent development is the requirement of standardized test scores to advance to upper level courses (Mingle, 1985). In Mississippi, the ACT-COMP must be taken in order to enter the teacher education program. Florida requires students to pass a minimum competency exam in order to advance to the last two years of coursework. In Georgia, students may begin taking the Regents tests as sophomores and may retake it as many times as necessary. It must be passed in order to graduate. Beginning in 1986, students at The University of South Dakota must take the ACT-COMP in order to enter teacher education.

Thus, grades continue to be the valuable that faculty use to measure student achievement in their courses. This valuable is being questioned today, and is increasingly being replaced by standardized tests taken by the students at the end of their sophomore and senior years. This seems to imply a lack of confidence on the part of the public, corporations and college administrators in faculty grading of their students.

Exit Evaluation

Until recently, little attention has been given to measuring student performance at graduation. A 1978 survey of institutions involved in accreditation related self-studies found that only one in

Figure 3. Increase in average intercorrelation as the percent of plus or minus grades increased



three had generated or examined data on student growth and learning (NIE, 1984). Another finding was that only 23% had measured students' knowledge in their major field.

This situation is changing rapidly, as states increasingly mandate the use of standardized tests for judging student achievement at graduation in an effort to tighten exit standards. Like other states, such as Tennessee, skills of South Dakota seniors will be evaluated and then compared to other evaluation data to measure the value added. Other states (e.g., Georgia) have identified licensing exams as exit measures, especially in the teaching and nursing areas (Mingle, 1985). The Tennessee legislature is requiring institutions of higher education to quantify evidence of improvement in student standardized scores at graduation (Mingle, 1985).

Studies of student performance on subject-area tests of the GRE and other standardized tests have been conducted recently. From 1964 to 1982, student performance on 11 of 15 major subject area tests of the Graduate Record Exam declined (NIE, 1984). A follow-up study of student performance on 23 different standardized tests during the same time found declines in 65% of the tests (Mingle, 1985).

In the midst of increased testing activity, the NIE Study Group cautions that evaluation should be focused on measuring learning at the end of a bachelor's program (NIE, 1984). In order to do this, they recommend the use of a systematic program of assessment and feedback of student knowledge, capacities and skills. The Study Group recommends a shift from "input" data which describes the entering student and the institution's resources as well as measures the competence of students at the end of a course, to output data which measures the growth of students from entry to graduation. In 1983, the Southern Association of Colleges and Schools discussed, but did not adopt, a proposal requiring its member institutions to evaluate students systematically and to focus on outcome measures (Mingle, 1985).

Thus, exit evaluation is becoming an important aspect of judging student achievement in an undergraduate program. At the same time that states are moving in this direction, however, the NIE Study Group cautions that we must keep our sights on student learning, not just on student test scores.

Summary and Recommendations

The concern about measuring what a student has actually learned versus assigning grades is an historical one. The concern was voiced in the late 1940's and early 1950's at a time when enrollments expanded and a national movement to improve college and university teaching developed. This same concern was also expressed in the 1960's and in the 1970's. Today concern is also expressed about learning versus earning credit hours (NIE, 1984).

From 1964 to 1974, grade inflation was experienced nation-wide. Since that time, the average GPA has remained fairly stable, indicating

that we continue to have an inflated grading system. Mingle (1985) identifies the loss of confidence in grading practices of faculty as one reason for state bodies to become involved in higher education. The NIE Study Group (1984) writes that colleges and universities "should establish and maintain high standards of student and institutional performance" (p. 3), implying that these standards do not exist universally now.

State initiatives into areas traditionally reserved for university faculty have increased dramatically in the 1980's and will continue to do so for several years. Involvement by state legislatures and Boards of Regents has probably also resulted from lack of confidence in higher education institutions' ability to maintain standards (Mingle, 1985). Certainly this involvement is an extension of earlier involvement for reforms in the K-12 grades.

One major result of state level involvement has been an increase in the use of standardized tests. While standardized tests were often required for college admission in the past, current practices are often at three data points: entry into college, progression into upper level undergraduate courses or programs, and graduation. The use of standardized tests allow for norm comparisons of large numbers of students across colleges and universities nationally.

The trend toward standardized testing is not receiving unconditional endorsement. A caution has been expressed by the NIE Study Group (1984) that test scores do not become the substitute for measuring learning. It recommends that a comprehensive approach to evaluating student learning be developed, with attention to graduation standards.

In light of these trends, the authors offer several recommendations. These are focused on the issues of grading and grade inflation, the use of standardized tests and the involvement of state agencies.

1. Colleges and universities should continue to moniter grade inflation at the course, discipline, department, school and college levels to identify and determine the possible causes for increases in grade point averages.

2. Colleges and universities should consider changing to a thirteen-point grading scale if the five-point grading scale (A to F) is currently in use. The thirteen-point grading scale includes plus and minus grades in addition to the letter grades used in the five-point scale. Also, it maintains the numerical values of the five-point scale. Its advantage is realized when grade inflation exists, as it currently does in most colleges and universities throughout the country. With increased grade inflation, there are fewer grade categories and therefore a decrease in discrimination, which affects the reliability (and in turn its validity) of the grade point averages. With the use of the thirteen-point scale, discrimination is increased by increasing the number of categories. As a

result the reliability and validity of the grade point average is also increased.

3. Colleges and universities should consider the use of criterion reference grading (CRG) rather than norm reference grading (NRG). CRG uses the same letter grades as NRG with the exception that the grades assigned to students to reflect their level of performance is judged upon their own merit with respect to some standards set by the instructor. This could be a way to begin to measure knowledge and skills as opposed to test-taking skills.

4. State agencies should involve university and college faculty in studying and adopting changes. Faculty morale is a key factor in maintaining and increasing student learning (NIE, 1984). Faculty and administrators should set output goals, such as students should be able to think critically, recognize cultural diversity or develop creatively. Data are already available about students at entry which should be used for counseling and placement in courses. As more data become available at exit, faculty and administrators should be involved in assessing to what extent students are meeting the goals of the institution.

5. Standardized tests should be developed at the state or local level depending on the financial feasibility. Since the United States does not have a national curriculum, the use of a series of national tests raises questions of validity. A study should be conducted of the influence of standardized tests on local curricula.

6. An instructional model which if based on systemmatic evaluation, such as that described by Gronlund (1985), should be used in identifying learning outcomes expected of students, preassessing students at entry, providing assessment feedback during the undergraduate program, and measuring for intended outcomes at graduation.

None of these recommendations can be easily implemented. However, each responds to a trend or concern that has developed over the past three decades. Reforms will continue while public concern is focused on higher education. Change should result from a cooperative approach to improving the credibility of grades and measuring learning.

References

Barnes, L. R. (1984). The negotiation of grades as a central feature of an educational exchange program. College and University, 59(2), 136-149.

Becker, H. S., Geer, B., & Hughes, E. (1968). Making the grade. New York: John Wiley and Sons, Inc.

Bejar, I. I., & Bleu, E. O. (1981). Grade inflation and the validity of the Scholastic Aptitude Test. American Educational Research Journal, 18(2), 143-156.

Donohue, W. R., & Struckman-Johnson, D. (1985, October). The value added of a college education. Understanding Undergraduate Education. Symposium conducted by The School of Education, University of South Dakota, Vermillion, South Dakota.

Gronlund, N. E. (1981). Measurement and evaluation in teaching. New York: Macmillan.

Hambleton, R. K., & Murray, John (1977, November). A comparative study and student attitudes toward a variety of college grading purposes and practices. Research in Education.

Komorita, S. S., & Graham, W. K. (1965). Number of scale points and reliability of scales. Educational and Psychological Measurement, 25, 987-995.

Kozak, R. (1985). Grade distribution analysis for The University of South Dakota. Vermillion, SD: University of South Dakota.

Masters, J. R. (1974). The relationship between number of response categories and reliability of likert-type questionnaires. Journal of Educational Measurement, 11, 49-52.

Millman, J., Slovacek, S., Kulick, E., & Mitchell, K. (1983). Does grade inflation affect the reliability of grades? Research in Higher Education, 19, 423-429.

Mingle, J. R. (1985). Measuring the educational achievement of undergraduates: State and national developments. Denver, CO: State Higher Education Executive Officers Association. (ERIC Document Reproduction Service No. ED 254 139).

Mortimer, K. P., Astin, A. W., Blake, J. H., Bowen, H. R., Gamson, Z.T., Hodgkinson, H. L., & Lee, B. (1984). Involvement in learning: Realizing the potential of American higher education. Washington, DC: National Institute of Education.

Oldenquist, A. (1983, May). The decline of American education in the 60's and 70's. American Education, 12-18.

Suslow, S. (1977, March). Grade inflation: End of a trend? Change.

Suslow, S. (1976). A report on an interinstitutional survey of undergraduate scholastic grading, 1960's to 1970's. Berkeley, CA: University of California, Office of Institutional Research. (ERIC Document Reproduction Service No. ED 129 187).

White, F. M. (1975, February 3). Sometimes an A is really an F. Chronicle of Higher Education, 9, 24.