

DOCUMENT RESUME

ED 047 164

AA 000 657

AUTHOR Warren, Jonathan R.
TITLE College Grading Practices: An Overview
INSTITUTION ERIC Clearinghouse on Higher Education, Washington, D.C.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
REPORT NO R-9
PUB DATE Mar 71
NOTE 29p.

EDRS PRICE MF-\$0.65 HC-\$3.29
DESCRIPTORS Academic Performance, *Evaluation Techniques, Grade Point Average, Grade Prediction, *Grading, *Higher Education, *Measurement Techniques, Pass Fail Grading, Recordkeeping, Student Attitudes, *Student Evaluation

ABSTRACT

This report examines grading practices, the uses of grades and the influence of grades on the student, faculty, administration and society. The author also indicates how grading practices can be and are being altered to provide an educational tool that accurately reflects the many dimensions of student performance. It is noted that grades seem unnecessary for many of the administrative purposes within an institution, other than as an indicator that a certain course has been passed by a student. Furthermore, selection for academic awards, honor programs or special classes could be based on faculty nominations supplemented by evaluative information provided by the faculty. It is concluded that more varied and effective grading procedures are available; however, that they are seldom employed may be caused by the uncertainty over what is really wanted of grades. In light of this, the components and structure of grades need closer scrutiny so that the issues raised by the grading process--involving, as they do, all levels of society--can be dealt with. (WVM)

ED 047 168

College Grading Practices: An Overview

REPORT 9

AA 000 657

COLLEGE GRADING PRACTICES: AN OVERVIEW

Jonathan R. Warren

Repo:19

**ERIC Clearinghouse on Higher Education
The George Washington University
1 Dupont Circle, Suite 630
Washington, D.C. 20036
March 1971**

FOREWORD

In this comprehensive review of the literature, Jonathan Warren examines grading practices, the uses of grades and the influence of grades on the student, faculty, administration and society. He also notes the relationship of grades to the social structure and cites a need for a clear definition of the purposes of grades. The author, a Research Psychologist with the Educational Testing Service, indicates how grading practices can be and are being altered to provide an educational tool that accurately reflects the many dimensions of student performance.

The ninth in a series of reports on various aspects of higher education, this paper represents one of several kinds of Clearinghouse publications. Others include short reviews, bibliographies, and compendia based on recent significant documents found both in and outside the ERIC collection. In addition, the current research literature of higher education is abstracted and indexed for publication in the U.S. Office of Education's monthly volume, *Research in Education*. Readers who wish to order ERIC documents cited in the bibliography should write to the ERIC Document Reproduction Service, Leasco Information Products, Inc., 4827 Rugby Avenue, Bethesda, Maryland 20014. When ordering, please specify the ERIC document (ED) number. Payment for microfiche (MF) or hard/photo copies (HC) must accompany orders of less than \$10.00. All orders must be in writing.

Carl J. Lange, *Director*
ERIC Clearinghouse on Higher Education
March 1971

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. GRADING EFFECTIVENESS	2
III. GRADING SYSTEMS	4
IV. PURPOSES OF GRADES	8
V. UNINTENDED EFFECTS OF GRADES	16
VI. TECHNICAL ISSUES IN GRADING	20
VII. POSSIBLE DIRECTIONS FOR COLLEGE GRADING	24

This publication was prepared pursuant to a contract with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of views or opinions do not, therefore, necessarily represent official Office of Education position or policy.

I. INTRODUCTION

In 1964, a report of a conference on grades in higher education emphasized that grading was perennially overlooked when the processes of higher education were considered. The conferees pinpointed some neglected but important questions about grading, such as (Teaf, 1964):

- Do we know the effects of grades on the educational process?
- Can grades be justified as incentives to learning?
- What aspects of student behavior are reflected by grades?
- What function do grades serve in selection procedures?
- Can alternative devices serve the functions of grades and eliminate their shortcomings?

Grading has slowly emerged from an area of neglect to become a widely discussed, controversial topic. But focus on grades, though intense, has been haphazard, in that only one or two of the issues raised in 1964 have been examined while other, more inclusive questions have been almost totally ignored.

More than one reader interpreted the review as a biased appeal for the abandonment of grades. The review is, of course, biased, but the favored position is that grading practices should be improved to make them serve their various intended purposes more effectively—not that they should be abandoned. This position implies that purposes should be identified and the effectiveness of grading procedures in serving those purposes should be compared with that of alternative procedures. The value of accomplishing the purposes should also be weighed against the cost of grading in terms of the expenditure of educational resources and of whatever undesirable side effects can be demonstrated. Desirable side effects should be weighed in favor of grading. The desirability or undesirability of side effects, however, is itself likely to be a disputed issue and would also merit study.

An initial bias to the effect that grading procedures could be improved has been immensely strengthened as the literature was read and evaluated. Present grading procedures are monolithic at the same time that higher education is increasing in diversity and complexity. Techniques of information processing and management, incorporating the basic functions of grades, are growing in power, subtlety, and refinement while grading processes remain at a standstill. The biases that pervade the following

pages, therefore, are that present grading practices are inadequate to their intended tasks; that possibilities for improvement are enormous and should be pursued; that purposes as well as practices require examination; and that the social as well as educational effects of grading are too important to be neglected any longer.

Of approximately 200 articles, papers, and reports about grades appearing from 1965 to 1970, about one-fourth considered the form of grades, especially whether Pass-Fail should replace A through F. Another one-fourth considered the use of undergraduate grades to predict grades in graduate and professional schools. Therefore, one-half of the recent literature on grading was occupied with only two limited aspects of grades—their external form and their predictive relationship to later grades. The remaining half of the literature ruminated over a variety of topics—variability in grading standards, disadvantages of grades, effects of grades on students, use of grades to predict occupational success, determinants of grades, and the social effects of grades—none of which appeared in as many as 10 percent of the total publications. (Excluded from this count is the large number of articles on the prediction of undergraduate grades.)

These reports, in spite of their variety, leave large gaps in our knowledge about grades and grading. They lead to only a few general statements that can be made with much confidence: students approve of Pass-Fail grading, but when offered a Pass-Fail option, they often don't elect the option to take courses they otherwise would not have taken; deans and others concerned with admission to graduate and professional schools disapprove of Pass-Fail grading in undergraduate colleges; undergraduate grades predict first-year graduate and professional school grades about as well as they have for years but not very well most of the time, occasionally quite well, occasionally not at all.

These results do not constitute an impressive advance in knowledge about an important, ubiquitous process in higher education. Still neglected, except in occasional speculative musings, are questions about the purposes of grades. For example: Are the purposes worthwhile? If so, are they well served? Are the frequent criticisms of grades justified? If so, can ways be found to serve the purposes of grades without the deficiencies of present procedures? While experiments with Pass-Fail procedures and prediction studies touch on parts of these questions, the basic issues remain obscured.

II. GRADING EFFECTIVENESS

Grades can be defined as sets of symbols that represent a level of academic achievement indicated by some form of evaluation. Their purpose is to condense the results of evaluation into a form simple enough for a continuing and cumulative record of student accomplishment to be maintained. The grading process is therefore not the evaluation process, but follows it. The effectiveness of a grading system can be examined with respect to (1) the fidelity with which it encodes evaluation results, (2) the ease with which it lends itself to recordkeeping, and (3) the adequacy of the information it conveys for the users of grades.

Fidelity

The translation of evaluation results into symbols is the most critical process in a grading system. Unless enough useful information is encoded by the grade symbol, effective functioning in other respects is almost worthless. As long as grades consist of a single symbol assigned for each course, they can convey information only on a single dimension, although several different kinds of performance might be observed in the evaluations on which the grade is based.

The term most commonly applied to the complex dimension grades are intended to measure is academic achievement. Yet academic achievement is itself defined only in terms of composition of course grades. It has no independent definition against which the validity of course grades can be checked.

The poor fidelity of grade symbols is largely responsible for the sparseness of the meaning in academic achievement (Erickson, 1966; Trow, 1968). The grading process begins with an individual instructor evaluating a variety of student performances—responses to test questions; the quality of thinking, understanding, grasp of factual detail, integrative ability, and fluency of expression found in written papers; the evidence of student preparation, understanding, and interest revealed in class discussions; and whatever other kinds of evidence the instructor considers relevant to his definition of achievement in that course (a definition that is probably unspecified). Indicators of all these components of achievement are then weighted and combined into a single scale often inappropriately because of differences in the variances of the indicators (Lacey, 1963). The composite measure is reasonably reliable, with respect both to internal consistency and test-retest reliability over periods shorter than a year. Grades may therefore be accurate in reflecting performance on some undefined dimension of academic achievement. But their fidelity is poor in that they transmit only a small part of the information in the evaluations that led to the grade while leaving the information they do transmit difficult to interpret.

Recordkeeping

Recordkeeping is facilitated by dividing achievement into some arbitrary number of segments. The number can range from two to more than 100 (some military institutions using a 400-point scale). Dressel and Nelson (1961) noted the hardness of the five-category scheme for segmenting the achievement continuum, pointing out that departures toward more or fewer categories ultimately revert to five. Two- and three-category schemes are either modified by pluses and minuses or otherwise subdivided, and schemes of more than five tend to have categories merged.

The division of the achievement continuum into segments or grade categories and the location on the continuum of the boundaries between categories and of instances of student performance are problems that have absorbed much attention. Comparisons of departments within institutions and of faculty members within departments as to their choices of location for the boundaries between grade categories are common (e.g., Juola, 1968). Attempts to find some reasonably stable common standard on which to anchor the achievement dimension are less frequent but still common (Anderhalter, 1962; Berdie, 1965; Fricke, 1965; Grant, 1956). Currently the number-of-categories issue—whether five or two—is being vigorously debated.

The convenience of a limited number of categories, or the difficulty in using more than about five, probably accounts for the strong tendency noted by Dressel and Nelson (1961) to reduce larger numbers of categories despite demonstrations that using fewer categories necessarily decreases grading accuracy (Ebel, 1969). The common practice of placing the boundaries between grades at points on test score distributions (or on distributions of accumulated points in a course) where breaks occur between adjacent scores is another accommodation to convenience. No justification can be found for assuming that gaps in score distributions have any relationship at all to what are presumed to be commonly accepted categories of performance. For convenience, something that occurs by happenstance is used to define the boundaries between grade categories.

Kirby (1962) pointed out that at one "rather large institution of good reputation" discontinuities at the boundaries between grade categories can be expected to cause 42 percent of the students to gain or lose, relative to their precise position on the achievement dimension, one grade point or more in 15 units of class. One percent of the students will gain or lose five grade points or 0.33 points in their grade point average for that semester solely because of errors due to discontinuity between grade categories.

Purposes of grade information

The users of grade records seldom seem concerned about the nature of the information conveyed. Yet the adequacy of the information encoded in grades is dependent on the purposes for which grade records are used. Different purposes might reasonably be assumed to call for different kinds of information, and purposes therefore merit examination. Is the substantial expense borne by an institution to maintain grade records justified to provide personnel evaluation and selection services to other agencies (Goodman, 1964; Jencks and Riesman, 1968)? What are the internal purposes for which grade records are maintained? Could several simpler and collectively less expensive procedures serve the same purposes? For example, should procedures for advising students about course selection and procedures for determining eligibility for extracurricular activities depend on the same set of records? Would the separation of records for different functions improve the effectiveness of each?

Eligibility for veterans' benefits, retention of scholarship awards, and draft status have depended on a student maintaining satisfactory academic standing. The kind of information required for these purposes differs substantially from that required by a graduate department selecting 10 students from among 50 applicants. Yet both kinds of purposes now depend on the same source of information, even though the graduate department may supplement the overall gradepoint average with other information. The primary information in grade records is still some kind of weighted average of the one-dimensional course grades. Each grade is a composite of a number of varied kinds of judgments, each composite differing in some unknown way from the others. Then these poorly defined composites are averaged into something that can only represent whatever does not distinguish a good memory from depth of understanding, or sensitivity to professors' preferences from imaginative synthesizing of disparate elements, or problem-solving ability from expository fluency.

The information in gradepoint averages may or may not be adequate for its purposes, but it can be no better than the information encoded in the original grades and is, in fact, substantially less than the total information in the collection of course grades. Deans and admissions officers who object to two-level, Pass-Fail grading on the ground that they need the greater amount of information in five-level, A through F grading systems are partly deluding themselves, for the additional information in five-level as opposed to two-level grading is almost uninterpretable. A greater number of grade categories does carry more information in a technical sense; differentiation among students is more accurate with five grade levels than with two (Ebel, 1969). But interpreting that increment in information—recovering from it the meaning that was in the original evaluation—is essentially impossible. The meaning is lost in translating a variety of evaluations into one dimension of achievement and then averaging performances

on a number of these essentially undefined but probably quite different dimensions into a single index. Consequently, gradepoint averages are reliable measures of an undefined entity.

What the achievement dimension represents is ignored in most of the controversy over the desired characteristics of the scale used to indicate different regions on the dimension. However accurately different levels of achievement can be located on a scale, the symbol assigned to a point on the scale can include no more information than can be represented by that single dimension. The processes by which dissimilar kinds of performance are collapsed onto a single dimension, and even what those dissimilar kinds of performance may be, are lost in the concern over how many categories to break the dimension into and how to assign students to those categories. The fidelity—not the accuracy—of the translation of evaluations into grade symbols is therefore one of the most critical issues in grading, and one of the most neglected.

Evaluation and grading as distinct processes

The preceding discussion distinguished grading from evaluation. Grades are the symbols that formally indicate a student's general level of academic performance. Evaluation consists of the variety of processes—reading papers, giving quizzes, lab exercises, and exams, asking questions, listening to discussions, observing the quality of student questions—by which faculty members arrive at judgments about student accomplishment.

The failure to distinguish between evaluation and grading, or the assumption that the two processes are one, frequently leads to fruitless debate. Faculty members have spoken against reducing the number of categories in a grading system because they believe evaluation of student performance would be hampered. Yet faculty evaluation of student performance and the communication of its results to students can be carried out with no reference whatever to grades. The institutional demand that grades be assigned may force some instructors to evaluate students even if they see no need to do so. But grades in no way preclude evaluation, whatever their form.

Evaluation used primarily to improve student performance by serving a feedback function, by informing students of their progress while performance is still fluid, still being developed, has been termed *formative* evaluation. *Summative*, or terminal evaluation, in contrast, is intended to provide an appraisal of the final level of performance at the end of some period of instruction or at some point of discontinuity, more or less arbitrarily defined, as when a student has completed 15 weeks of instruction (Scriven, 1967). Grading is usually associated with summative evaluation, which often requires a different set of procedures to be most effective than does formative evaluation (Bloom, 1968; Husek, 1969). Summative evaluation and grading

may also be carried out by an agency other than the one providing the instruction. Formative evaluation, being part

of the instructional process, must stay within the control of the instructor.

III. GRADING SYSTEMS

Pass-Fail grading

The primary controversy over grading at present is whether multilevel grading systems, usually the five-level A through F system, should be replaced with a two-level Pass-Fail system. Distinctions among several levels of acceptable or adequate performance and one failing level would be replaced with the single distinction between acceptable and unacceptable performance.

Although some form of Pass-Fail grading has become common in the last 5 years (Benson, 1969; Buchman, 1970; Burwen, 1970; Quann, 1970), only a handful of colleges have put their entire grading system on a Pass-Fail basis. The typical procedure is to offer students an option of taking a limited number of Pass-Fail courses with the rest graded on the standard A through F basis. Dartmouth's option procedure is representative of most. Students were permitted to take one Pass-Fail course per term provided the course was not in the student's major field (Feldmesser, 1969). Other colleges limit the option to seniors, upperclassmen, or those with grade point averages above some minimum. Courses in the student's major field are almost always excluded, while courses needed to satisfy foreign language or mathematics requirements are sometimes and sometimes are not excluded. More than one Pass-Fail course per term is seldom permitted. With these limitations, few students complete college with more than 10 percent of their grades Pass-Fail (Pass-Fail Study Committee, 1969). The consequences of Pass-Fail grading, undertaken with great trepidation and concern, have been trivial.

The most common reason for adopting a Pass-Fail option is to encourage students to take courses they otherwise might not risk for fear of jeopardizing their grade point average (Benson, 1969; Feldmesser, 1969; Freeman, 1969; Johanson, Rossmann, and Sandell, 1970; Melville and Stamm, 1967; Milton, 1967; Morishima and Micek, 1970; Quann, 1970; Sgan, 1969; Stallings, Smock, and Leslie, 1968; Wharton, 1969). Students were expected to feel freer to explore unknown areas and to try courses in which they feel some insecurity. However, they have not used the Pass-Fail option for this purpose to any great extent.

At each of five institutions (Dartmouth, Princeton, Wellesley, the University of Michigan, and the University of Washington) where students were surveyed after initiation of a Pass-Fail option, roughly 75 to 85 percent of the students who elected to take a Pass-Fail course said they would have taken the course anyway (Cromer, 1969; Feldmesser, 1969; Karlins, 1969; Morishima and Micek,

1970). At Brandeis the pattern of course enrollment showed little change after a Pass-Fail option was instituted (Sgan, 1969). Some students apparently do take courses they would not take other than under a Pass-Fail option, but the number is not large.

Another common reason for adopting Pass-Fail grading procedures is to reduce student anxiety over grades. When asked, students have reported feeling less anxious in Pass-Fail courses (Cromer, 1969; Karlins, 1969; Melville and Stamm, 1967). In this respect the Pass-Fail option seems successful, although retrospective reports about emotional responses are typically not reliable indicators of actual responses. More dependable are the student reports that say overwhelmingly (but not unanimously) they like Pass-Fail grading. Students when surveyed inevitably urge continuation and expansion of limited Pass-Fail option procedures (Cromer, 1969; Erickson, 1967; Karlins, 1969; Melville and Stamm, 1967; Milton, 1967; Morishima and Micek, 1970; Priest, 1969).

Other reasons given for Pass-Fail grading are to shift students' efforts from grade-getting to learning (Benson, 1969; Committee on Educational Policy, 1970; Feldmesser, 1969; Milton, 1967; Quann, 1970; Sgan, 1969); to let the teacher function as mentor rather than judge (Committee on Educational Policy, 1970); to avoid the preense that students are evaluated more accurately than is the case (Benson, 1969); and to give students greater control over the allocation of study time (Milton, 1967). While these seem plausible expectations to hold for Pass-Fail grading, only the last can be supported by evidence (Erickson, 1967; Feldmesser, 1969; Freeman, 1969; Karlins, 1969; Morishima and Micek, 1970).

The tendency of students to slight courses graded Pass-Fail in order to concentrate on other courses has been offered as a defect in Pass-Fail options. Yet the view that student control over their distribution of effort is desirable seems more defensible (Milton, 1967). A course may have a particular interest or be particularly important to a student's major field or be more difficult for him than others. These all seem good reasons for students to adjust their effort unevenly across different courses. Elton (1968) and Feldmesser (1969) have used similar arguments to propose schemes for variable weighting of course grades with the students choosing the weights to be assigned.

One might speculate that what some faculty members object to is not the differential allocation of effort to different courses as much as the possibility that students may go through college, or at least through some courses,

without expending an acceptable amount of effort. Instructors who use grades as a device for coercing students into kinds of behavior the instructor considers desirable (Mayhew, 1969) or who adjust their grades according to the amount of effort the students are believed to have expended (Axelrod, 1964) might be expected to feel chagrined when students manage to learn without going through the tasks set by the instructor. This view cannot be advanced on the basis of clearcut evidence; its plausibility can only be inferred from unsystematic observations and experience and the general expectation that people with hostile, punitive proclivities can be found among college professors as well as elsewhere.

The major objection to Pass-Fail grading is the problem of graduate and professional school admission. However, that objection is serious only if a substantial part of a student's record consists of Pass-Fail grades, something that occurs in only a few colleges. At the University of California at Santa Cruz, one of the few institutions where most grades are Pass-Fail, more than half of the graduate school aspirants among the 1969 graduates reported they encountered no problems in gaining admission. Nine percent did report problems and another 35 percent were not sure (Pitcher and Bosler, 1970). Although the Pass-Fail grading system had affected graduate school selection to some extent, most students who applied were admitted, although not always to the school that was their first choice. Perhaps more serious than not attending a first-choice graduate school was the loss of fellowships as a consequence of the Pass-Fail transcript. This did occur but its frequency is not known.

Whitman College reverted to the customary A through F system after 15 years of Pass-Fail grading primarily because of difficulties encountered by student transfers and by graduates applying to graduate schools (Perry, 1968). Yet perhaps because of the growth in concern over grading since Whitman's abandonment of Pass-Fail, the difficulties encountered by Santa Cruz graduates were not considered great enough to induce a similar action there (Committee on Educational Policy, 1970). The prestige of the undergraduate institution may also affect graduate admission, although Whitman's difficulties occurred in spite of a strong academic reputation.

Pass/No Record grading

A system similar to Pass-Fail grading has been proposed in which failure results in removal of the course from a student's record. The primary argument for Pass/No Record grading is that failure to achieve an adequate level of performance in a course should not result in a penalty to the student. He should simply not be given credit for the course. Brown University has instituted such a procedure and Stanford is considering a recommendation to do so (Laid, 1970). Several other institutions have either tried this system, on an experimental basis or are considering it

(Christin, 1970; Goldstein and Tilker, 1969; Smith, 1969).

A number of junior colleges have instituted some form of this "nonpunitive" grading. All D's and F's are replaced by W's, which indicate only that the student should not be given credit for completing the course. Recordkeeping practices differ. At Santa Fe Junior College in Florida the failed course does not appear on the student's record (Fordyce, 1969-70), making it a Pass/No Record system ("Pass-Erase" in the Stanford terminology). Other junior colleges record the "No Credit" grade, which means that a record of attempting a course and not passing it is maintained (Brooks, 1968; Smith, 1969); however, the course is not counted when a student's grade point average is computed. The failed course can be repeated as often as the student chooses until he passes.

However, many colleges and universities, and probably most, require a minimum grade point average for a student to be readmitted the following academic year. This means that standards are raised for the second academic term relative to the amount a student has fallen below the acceptable first-term level. Making the hurdles higher as the performance level drops seems an unreasonable procedure. Pass/No Credit grading avoids that situation.

Some faculty members object to Pass/No Record grading because a student could stretch out indefinitely the time he spends accumulating enough units to graduate. At colleges having a student body homogeneous in previous preparation and aptitude this might be a valid objection. On the other hand, though, in a homogeneous student body the number of students stretching their time in college inordinately would probably be small. The basic argument is whether students taking courses in which they can fail without penalty would constitute an inefficient use of the institution's resources. No one knows.

At junior colleges, where substantial proportions of entering students have not been successful in previous educational settings, early demands for a uniform level of performance seem particularly questionable. Many junior college entrants need a period of adaptation to college, and the Pass/No Credit system allows this to them. At more selective institutions like Stanford and Brown, the same opportunity for adaptation might be desirable if heterogeneity in the student body were to be increased. In any case, though, whether students are to be permitted to move through college at varying rates is a question to be decided on its own merits. That decision should then enter into consideration of whether or not to record unsatisfactory performance in a course.

Marshall (1968) has described in some detail the process by which the faculty in a department of a medical school reached a decision about grades. After extended discussion of various procedures, one faculty member observed that in the particular situation in their department the most useful distinction to be made with respect to student performance was between students who had clearly mastered the content of a course and those about whom there was some question. The most sensible grading scheme, and the one

that was then adopted, was Clear/Not Clear with respect to evidence of course mastery. In most graduate and professional schools and in selective undergraduate colleges, the number of students who are clearly incapable of mastery at an acceptable level of some program of courses is likely to be quite small. Whether or not Clear/Not Clear would be the most appropriate grading scheme at all such institutions, as it seemed to be at the University of California Medical Center, consideration of the nature and purposes of grades at that institution should underlie the selection of a grading system.

Descriptive grading

Descriptive grading, which historically preceded the various symbolic grading scales being used (Smallwood, 1935), consists of written comments that describe the student's performance. It is not based directly on any scale of academic performance except the implicit and intuitive scales that underlie a professor's judgments. Both the nature and the level of the performance are described by the instructor, and both may vary within a single class or course. This specification of the nature of the performance that leads to a judgment of excellent, good, or poor is the major distinction between descriptive grading and symbolic grading. With symbolic grading scales, differences that commonly exist in the nature of the performance evaluated are lost.

The most serious drawbacks to descriptive grading are the time required for faculty members to write the descriptive comments and the difficulty in making quick and simple comparisons of performance descriptions. Forming a judgment about which letter grade to assign a student is often, though not always, easier and less time-consuming than writing a descriptive comment about a student's performance. Comparing the capabilities of two students is also easier if each is described by a numerical gradepoint average instead of a set of instructor comments. But while the process of comparing is simpler with gradepoint averages, the information on which the comparisons are based is probably far greater with descriptive comments. Even a terse, relatively barren comment, such as "good student," that opponents of descriptive grading point to as illustrative of its weakness, is at least as informative as "B."

The most detailed accounts of long-term experiences with descriptive grading are those of Sarah Lawrence College (Murphy and Raushenbush, 1960) and the microbiology department of the University of California Medical Center (Marshall, 1968). The University of California at Santa Cruz has used a combination of Pass-Fail and descriptive grading since its opening in 1965 (Committee on Educational Policy, 1970), and a few other colleges, usually small, selective, liberal arts colleges such as Bennington and Goddard, have used descriptive grading or some combination of description and symbolic grades.

The strengths and weaknesses of descriptive grading are closely associated with the purposes for which grades are intended. Its major strength is in specifying the dimensions of the evaluated performance. If feedback to students can be accepted as a grading function, descriptive grading can be superior to other forms. It may not be superior if the descriptions are inaccurate, misleading, or uninformative, but its potential for conveying information is far greater than that of symbolic grading.

The weakest aspect of descriptive grading is its cumbersome nature for selection and other administrative processes involving large numbers of students. Yet modern information storage and retrieval techniques appear able to manage descriptive grading as effectively as symbolic grading has been managed in the past. Recording and storing prose descriptions of student performance seem feasible. From the stored descriptions, reports of student performance could be compiled to summarize only those elements relevant to the purposes for which the report is intended. Selection for employment and selection for graduate education are two purposes that might be expected to rely on evaluations of different kinds of student performance and therefore would require different reports.

Other grading systems

The grading procedure at the University of Surrey combines level of student performance with course difficulty, difficulty being determined by both course level and intensity (Elton, 1968). Student performance is judged in conventional ways from examinations, essays, projects, and other course work. The student's grade is then the product of his level of performance and the difficulty of the course.

A joint student-faculty committee at the University of California School of Law in Berkeley, after extensive interviews with students, faculty, alumni, and employers, recommended changes to the existing procedures to give them more flexibility and make them more informative (Committee on Grading, 1970). The existing system was a three-point scale (Top-Middle-Low) with 10 percent of the students in any class assigned to each of the extreme categories and 80 percent assigned to the middle category. Faculty and students objected to the rigid proportions in which grades were to be assigned and to the lack of differentiation within the middle category.

The Committee's recommendation was to use three levels of passing grades—Excellent, Very Good, and Qualified. Variable proportions of students can be assigned to each level, depending on the instructor's judgment of the overall performance of the class. From 15 to 20 percent of the students in a class would be graded Excellent, for example, and from 30 to 35 percent Very Good. The rest of the students who reach an adequate level of performance would pass as Qualified. Students who did not reach an acceptable level would receive an Incomplete, to be removed by repeating some or all of the course.

The proposed law school procedure is similar to the ABCX procedure of some junior colleges, where the X indicates inability to undertake the next course in a sequence but does not appear in the student's record, and can be removed by repeating the course if the student chooses. The law school system differs by specifying a range of proportions for each grade and in requiring that inadequate performance be brought to a satisfactory level.

A grading procedure that would allow for diversity in the kinds of student performance evaluated, and make the various kinds of performance explicit, has been proposed by Elbow (1969). He suggests that a list be provided of those aspects of student performance considered important by the faculty. The students in any particular course would then be graded with respect to those qualities listed that the instructor in that course considered pertinent.

The qualities rated would almost certainly differ across courses, and they could also differ within a class. Except in large classes, instructors commonly have different kinds of information about different students. Qualities observed in certain students and graded by the instructor may be left ungraded for other students in the same class because no occasion for their observation occurred.

The variables graded may also differ at various levels of performance. An instructor might consider diligence important in a low-performance student but not relevant to a high-performing student. Since creative integration of disparate elements into an effective construct may only appear among high-performing students, rating all students on that dimension would be unnecessary.

This procedure combines elements of descriptive and symbolic grading. The descriptive phrases or dimensions of performance are provided in advance, limiting the instructor's freedom of invention. Systematic determination of those qualities most often considered important by the faculty, however, could make this an unimportant consideration. For those who prefer the present one-dimensional grading scheme, one of the dimensions offered for rating might be general academic performance. The number of levels of each rating could be two or three or more, but a limited number seems preferable. Recordkeeping and reporting of grades would be somewhat more cumbersome than with single grades but would not be a serious problem.

Hoyt (1966, 1968) made a similar proposal in recommending that grading be multidimensional and reported in the form of a profile. The primary advantage of such a scheme is in specifying the nature of the performance evaluated and intended to be reflected by a grade. Averages would also be profiled to refer to specified kinds of performance. Persons using grades for selection would be able to make their own judgments about the kind of performance they consider important and would no longer have to assume that the evaluator had the same views of what constitutes desirable performance as the selector.

Scouts for professional football teams use a scheme such as this for grading college players. Six or eight dimensions

are provided on a form for grading players at each position. Some dimensions are specific to a single position while others are common to several positions, but distinctions among different kinds of performance are explicit. Offensive backs, for example, are graded separately on running with elusiveness and running with power but are not graded on receiving long passes. Offensive ends are graded separately on receiving long passes and receiving short passes but are not graded on power running. Both backs and ends are graded on "hands"—the sureness with which they handle a football (*San Francisco Sunday Examiner & Chronicle*, 1971). Academic performance is surely more complex than football and selection to graduate or professional school more important for society than selection of candidates for a job as a football player. But football selection is carried out with far greater discrimination.

A grading alternative that should not be ignored is abolishing all grades. This does not mean instructors would not evaluate student performance in whatever ways are appropriate or that the results of those evaluations would not be communicated to the students. But no formal record would be made of the level of a student's performance in a class. Records would only indicate satisfactory completion of a course.

Only one of the major purposes for which grades are intended would be jeopardized by their abolition. Other institutions would have to find other criteria for selection. In view of the erratic performance of grades in selection, however, this seems not to be a serious consequence. The greater looseness in selection procedures for graduate and professional training would probably complicate the tasks of admissions officers, but the social benefits of the increased heterogeneity of the population entering graduate and professional training might well justify the admissions officers' problems. A distinct benefit would be the forcing of graduate schools to give closer attention to the selection process and its purposes.

The motivational and informational functions intended for grades are questionably served if at all. The limited evidence available suggests that their motivational effects vary with different kinds of students in different kinds of situations and may not be great in comparison with other motivating forces, such as the desire to perform well. The informational function of grades is negligible as far as students are concerned if the results of the evaluation process are effectively communicated to them. The institution has little need for records of student performance level. The courses a student has completed satisfactorily are enough. Awarding academic honors and financial aid (if financial aid is to be based on level of performance) can be based on faculty nominations or other derivatives of faculty evaluations that would not require grades for all students.

In short, the abolition of grades is not an unthinkable alternative. It may turn out not to be desirable, depending on circumstances and the desirability of the purposes, but it

merits consideration. If grades do serve a useful function that justifies their cost to the institution, that fact should be established more surely than it has been.

A few colleges have apparently functioned well without grades in the past and continue to do so. They tend to be small, selective liberal arts colleges, medical schools, or

small experimental programs within a college. But the experiences of these institutions show that in some circumstances grades can be abolished without undesirable consequences. Education without grades presents problems, but these problems may be far less serious and more amenable to solution than the problems grades contribute.

IV. PURPOSES OF GRADES

Much of the preceding discussion, but particularly the distinction between formative and summative evaluation, suggests that evaluation and grading procedures have several purposes and should vary to accommodate different purposes. Yet the literature on grading almost totally ignores what purposes grades are intended to serve, except for the fairly frequent complaint that purposes are too often ignored (Dyer, 1967; Fricke, 1965; Milton, 1966; Korn, 1969; Scriven, 1969; Westland, 1969; Wolfe, 1968). Even Thorndike's presumably comprehensive review in the *Encyclopedia of Educational Research* (Thorndike, 1969) dealt primarily with the difficulties to be overcome if grading is to improve. Thorndike, and the body of research he reviewed, treated grades like the weather. They are an inevitable part of educational life, and the best we can do is accommodate them. Whether accommodation is preferable to their abandonment, or whether substantially different procedures might better accomplish the purposes for which grades are intended, are apparently seldom considered.

Grading, according to Scriven (1969), is a fundamental educational practice particularly in need of empirical investigation with respect to the purposes and values it serves. That such investigations have not been made is attributed to the reluctance of researchers to examine questions of social values or moral issues. Such questions are considered beyond the reach of empirical investigation. Yet the distinction between facts and values on which researchers base their avoidance of value-oriented research is spurious. Decisions involving questions of merit, worth, or value should have empirical justification.

Stake (1970), in discussing the evaluation of educational programs, urged that more attention be given to empirical studies of the goals and values that determine criteria of performance. His argument holds equally well for the evaluation of students. The purposes of evaluation and selection of the kinds of performance to be evaluated are issues amenable to empirical study.

The discussion of grading purposes that follows rests only indirectly on empirical data. Studies to guide the selection of purposes, to direct educational decisions that touch on social values or moral questions, have not been attempted. The competitive aspect of grading, for example, has been cited as both desirable and undesirable, yet very little evidence is available to support either view. Nevertheless, many discussions of grading practices start with an unexamined assertion that grades have a stated purpose.

Some reports of current practices clearly imply one or more purposes for grades. From these statements and implications, the generally accepted purposes of grades can be determined even if the justification for their acceptance cannot.

In one of the few general considerations of grading purposes, Erickson (1967) made two important distinctions. Grades can serve either administrative or educational functions and both functions, in turn, can serve either students or the institution and society at large. Grades provide a reasonably standard way of recording student progress and performance for administrative decisions about retention or dismissal, selection, transfer, honors, or extracurricular participation. Educationally, grades are intended to help students and professors alike to adjust their academic programs and activities to make the most effective. Although evaluation rather than grading usually accomplishes this function, this aim is often advanced as one of the important purposes of grades.

The administrative functions of grades usually serve institutions, while their educational functions serve students as well as institutions. Whether these different functions conflict, and if they do, how precedence is to be determined among them, are questions that should be probed through empirical studies.

Sorting and selecting students

By an overwhelming margin, the most commonly discussed purpose of grades is their use as a device for screening and selecting students for more advanced education, employment, fellowships and awards, honors, transfer to other institutions, and participation in institutional activities. This is an administrative rather than an educational function, and serves the institution or society rather than the student. Its disproportionate attention in the literature indicates a tacit assumption of priorities that justifies closer examination.

Glazer (1970a, 1970b) argued for the usefulness and importance of grades as a method of ordering students with respect to academic merit. Successive selection to higher educational programs on the basis of merit progressively differentiates the population with respect to academic accomplishment. This ensures that as selection becomes increasingly rigorous, the most capable people face the

most demanding tasks, at least in terms of academic performance. The resulting concentration of people with high academic capability further enhances their productivity. Dispensing some of the rewards of society in accordance with academic merit is highly defensible in view of the importance to society of the academically capable. Grades, as the mechanism by which people are sorted according to academic merit, are therefore quite important and are a more equitable mechanism for distributing society's rewards than is parental social status, which they to some extent replaced.

Jencks and Riesman (1968) gave a contrary interpretation. The academic achievement that grades reflect is a somewhat circumscribed kind of performance more readily attained by members of higher social and economic classes than by those of other classes. Yet education is also the primary path to higher social and economic status. Consequently, educational selection based on previous performance offers the opportunity for further development to those already most highly developed and increases the gap between the lower and upper segments of the population with respect to whatever benefits education provides. As those benefits become more strongly associated with power and prestige, formal education can be charged with exacerbating already serious social ills. Glazer's argument stressed that grades are essential primarily because they do differentiate according to academic performance and thereby make the distribution of social rewards more equitable than they otherwise would be. The critical point is whether academic performance is sufficiently important to be the basis for the distribution of large social rewards.

Sociologists and other critics of American education (Caplow, 1954; Friedenberg, 1970; Katz, 1968; Lauter and Howe, 1970; Sexton, 1967) have argued that one of its primary achievements has been to maintain the existing socioeconomic class structure, smoothing the way to socioeconomic advancement for those already possessing the desired social characteristics while systematically hindering and discouraging others. From the primary grades up, it is argued, those culturally unattuned to the dominant social class have been discouraged, shamed, and labeled incompetent. An important means for producing these effects has been the teacher assigned grade, which finds its justification in its consistency from teacher to teacher and from year to year. But consistency by itself has little to recommend it if the substance behind consistency is docility, compliance, agreeableness, and teacher-approved deportment instead of intellectual competence.

Whether this criticism is justified or not is difficult to determine. The view that the educational system maintains existing social inequities is based on subtle, long-term social effects that are altered by a variety of other social forces. Yet it is an enormously important issue that has been almost totally ignored in research if not in social comment.

A component of the educational process as pervasive as the system of grading and the resulting grade-based selection cannot fail to have important consequences for

society. That these consequences go beyond the training of a technically competent work force, perhaps in undesirable directions, seems probable. They deserve a kind of attention not provided by the current arguments over two-level versus five-level grading schemes.

Selection for advanced education

Graduate and professional schools are the primary beneficiaries of the use of grades for selection; they are also the group of "consumers" of grades most concerned about departures from traditional patterns (Dale, 1969; Hanlon, 1964; Hassler, 1969; Iadarola, 1969; Law School Admission Test Council, 1970; Rosser, 1970; Rossmann, 1970; Sparks, 1969). The deans of 230 graduate schools preferred overwhelmingly that their applicants present transcripts for evaluation that contain a predominance of letter grades. Yet a five-to-three majority indicated a reluctant acceptance of transcripts containing nothing but Passes if additional information about the applicant were available (Hassler, 1969). The Law School Admission Test Council, with representatives from almost every law school in the country, recently published a formal statement warning about the consequences to law school admissions of even partial Pass-Fail grading (Law School Admission Test Council (1970).

Typically, selection to graduate and professional schools is made from a pool of applicants much larger than the number to be admitted. Since most of the applicants are reasonably well qualified, distinguishing between the poorer of those admitted and the better of those rejected requires fine discrimination. Grade-point averages from undergraduate institutions, in spite of their deficiencies, permit such hairline distinctions, and the abundance of qualified applicants serves to keep selection errors low. Virtually all those selected are capable of acceptable performance. Errors of rejection may be more numerous but, by their nature, are seldom observed and present no problems for the institution.

The question of error in admission decisions highlights one of the problems in assessing the usefulness of grades in selection to higher educational institutions. For error to be measured, some definition of "correct" decisions is required. An admitted student who earns good grades and completes the course of study is considered to represent a "correct" admission decision. But deans and faculty members often deny that high probability of earning good grades is, by itself, an adequate basis for admission, and the correctness of decisions to reject applicants is neither defined nor measured. The usefulness of grades as a selection criterion cannot be adequately assessed until the purposes of selection are better defined.

An unexamined question in educational selection is whether an institution's educational resources should be denied to those not likely to receive high grades. Scriven (1969) indirectly raised this question by stating that one of

the three essential functions of grades is to provide "a basis for the allocation of scarce resources to those who can use them best (p. 114)." He did not attempt a definition of what constitutes the "best" use, and neither have others. The assumption that students who receive high grades have made better use of an institution's resources than have students who receive low grades may be justified. It has been questioned, however, and merits attention (Jencks and Riesman, 1968; Woodring, 1968). Furthermore, the *best* use of an institution's resources varies with the purposes of the institution. Medical schools, law schools, graduate schools of business, graduate schools of social work, and other institutions that award advanced professional degrees obviously differ in their purposes.

Even among institutions of the same type, purposes may differ. Some law schools, for example, consider their major function to be preparing students to pass the bar examination and enter legal practice. Others consider that purpose secondary to providing a legal education to all who might benefit from it, whether they become practicing attorneys or not. Still others place great importance on graduating those likely to produce advances in the present system of jurisprudence. Differences in educational purpose might imply different selection procedures—yet all depend heavily on undergraduate gradepoint averages.

The departments within a graduate school might also vary in their selection criteria. Gamson (1967), for example, showed that faculty members in the physical and social sciences differed consistently in their expectations for their students and in their own role perception in relation to students. (See also Riesman, Gusfield, and Gamson, 1970.) Yet studies have not been carried out that would allow selection procedures to be geared to different institutional or departmental purposes.

Anxiety over a possible threat to the selection function led to a survey of colleges having Phi Beta Kappa chapters for an assessment of the difficulties the growing use of Pass-Fail grading might present in electing students to membership. The Committee concluded that the use of Pass-Fail grading was not yet much of a threat to adequate evaluation of students because even when used it seldom constituted more than a fraction of a student's grades. The Committee further stated that grades should not be the only consideration in election to Phi Beta Kappa. In fact, two of the Committee's four recommendations urged de-emphasis of grades in election to Phi Beta Kappa (Pass-Fail Study Committee, 1969).

If the probability of earning good grades is accepted as the most justifiable basis for selection of students, problems still remain when gradepoint averages are used as a selection criterion. The effectiveness of previous grades as predictors of later grades has been examined extensively but with little depth, as is indicated by the large proportion of grade prediction studies that are doctoral dissertations. Experienced researchers with the resources to probe an issue deeply seem to find other problems more interesting.

For most students previous grades do predict later grades moderately well over relatively short time periods. Undergraduate grades predict first-year grades in graduate and professional schools moderately well, but they predict more advanced grades poorly, particularly in clinically oriented programs (Bartlett, 1967; Gough, 1967; Gohn, 1968; Hanlon, 1964). The number of studies that show negligible relationships between undergraduate and graduate school grades suggest that prediction is a rather selective process, operating differently for different people in different situations. If an institution should decide that the predicted gradepoint average is to be the dominant factor in deciding admission, the difficulty of predicting that average is still substantial.

The problem in depending heavily on undergraduate gradepoint average for selection to graduate and professional programs can be illustrated by showing the implications of a correlation coefficient of .30 between undergraduate gradepoint average and first-year graduate school grades. If the distributions of both grade averages are symmetrical and approximate a bell-shaped curve, a correlation coefficient of .30 will occur when 20 percent of the total group of students drop from the top half of the distribution of undergraduate grades to the bottom half of the graduate school grade distribution. Another 20 percent will move from the lower half with respect to undergraduate grades to the upper half with respect to graduate school grades. About 60 percent will achieve graduate school grades that put them in the same half of the grade distribution as did their undergraduate grades.

A sharper picture of a correlation coefficient of .30 can be seen by comparing the numbers of students who will shift their position with respect to grade quintiles. Only one-third of the students in the top 20 percent with respect to undergraduate grades will remain in the top 20 percent with respect to graduate school grades. Correspondingly, only one-third of those in the bottom fifth at entrance will remain there. Ten percent of the top one-fifth will drop all the way to the bottom fifth in graduate school, and ten percent of the bottom fifth with respect to undergraduate grades will move to the top fifth in graduate school. Greater numbers will move from the second to the fourth quintiles and from the fourth to the second. Clearly, a correlation coefficient of .30 would indicate a substantial amount of change in performance between college and graduate school.

The figure .30 has been chosen for illustration because it is close to the median value that has appeared in a large number of studies predicting first-year graduate school grades from undergraduate gradepoint average. Among about 40 studies reported since 1965, involving various kinds of graduate and professional schools and with several studies including from 10 to 25 different institutions, the correlations between undergraduate and first-year graduate school grades fluctuated rather widely on either side of .30. A report of the correlations obtained in graduate schools of business is illustrative. For the first-year classes in 1967-69

at 26 graduate schools of business, the median correlation between undergraduate and graduate school grades was .28 (Pitcher, Deemer, and Smith, 1968). For 19 of the same schools the mean correlation coefficients in 1954 and 1958 were .28 and .34, respectively (Pitcher and Winterbottom, 1965).

Klein and Evans (1968) reported correlations between undergraduate grades and first-year law school grades among seven law schools that ranged from .11 to .43 with a median of .33. Pitcher (1965) found similar relationships in 10 law schools for students entering in 1962, the correlations ranging from .10 to .39 with a median of .27. In a more recent report of students entering five law schools in 1966 (Schrader and Pitcher, 1970), four of the correlations between undergraduate grades and first-year law school grades were between .27 and .32. The fifth was .20.

Medical schools, dental schools, schools of social work, schools of education, and a school of veterinary medicine have shown similar results. Most of the individual correlation coefficients reported fall in the range from .10 to .50, clustering around .30 (e.g., Boldt, 1970; Bundy, 1968; Gough, 1967; Hepworth, 1969; Lunneborg and Lunneborg, 1966; Roemer, 1965).

In various graduate school departments the correlations between undergraduate and graduate school grades range somewhat more widely, from about .20 to .60 (Hackman, Wiggins, and Bass, 1970; Lannholm, 1968a; Lannholm, Marco, and Schrader, 1968; Mehrabian, 1969; Stordahl, 1967; Wiggins, Blackburn, and Hackman, 1969). In view of the great variability of the correlation coefficients and the fact that the extreme values tend to occur with samples of fewer than 100 students, little can be said with confidence about the relationship to be expected between undergraduate grades and graduate school performance. In selected circumstances the relationship may be quite strong, but what might produce those circumstances has not been identified.

The studies reviewed above, almost without exception, involved predictions of first-year graduate and professional school grades. Since predictions of second, third, and fourth-year grades can be expected to be successively lower, the utility of undergraduate grades as a device for making any but the grossest decisions about admission to graduate schools seems questionable. Since admission must continue to be selective as long as applicants far outnumber those who can be admitted, the alternative is to find more specific student attributes or combinations of attributes that are pertinent to the performance the selecting institution expects from its students. These attributes are not easy to specify; but until they are, selective admission will not be a very well developed process.

The weak relationships between undergraduate and graduate school grades can be excused on several grounds. Graduate students are a selected group; therefore, the distribution of undergraduate grades has been truncated. Graduate school grades also have a limited range, often only consisting of A's and B's. Yet pointing out reasons why a

predictive relationship is not high does nothing to improve the usefulness of the predictions. If graduate school grades cannot be predicted without substantial error, other criteria for selection should be sought, perhaps based on the particular purposes of the selecting institution or on other student characteristics desired by the faculty, such as those reported by Davis (1965) and Hilton, Kendall, and Sprecher (1970).

The heavy reliance on the grade point average in admission to advanced educational programs despite its questionable validity seems due to two factors. One is its administrative convenience. Since it is quantified, it has the appearance of accuracy and permits decisions based upon it to be objective. Decisions can then be made mechanically—which is often what is meant by objectivity. A comment from a respondent to a survey conducted by the Council of Graduate Schools (Hassler, 1969) illustrates this point: "Our Graduate School requires a 2.500 average on a 4.000 scale." What lies behind those three-decimal numbers remains unknown and unquestioned.

The second favorable aspect of the grade point average is its academic respectability. It reflects the combined judgments of a number of faculty members—people expected to make judgments from points of view similar to those of the faculty members in the selecting institution. It operates, therefore, like a set of recommendations to an exclusive club written by long-time members who know the kind of people the other club-members prefer. This is a harsh judgment and probably overstated. Generally, grades are the result of conscientious efforts at evaluation and of thoughtful, at times agonizing, decisions about grade assignments. They reflect the best judgments available about capabilities faculty members consider important. But the exclusive-club analogy again becomes appropriate, because no one can say just what kind of capabilities a faculty member had in mind when he evaluated his students and assigned grades. So grades and the grade point average are left with little more than their academic respectability vouched for by a member in good standing of the proper kind of club.

Parenthetically, the readiness of business firms—a different sort of club—to accept the recommendations of academic institutions is strange, particularly when recommendations of employers that say a person performed some business function very ably will have no influence at all in getting that person admitted to an academic institution.

The claims of validity for the grade point average and for its acceptability as the primary admission criterion rest on more than respectability, however. As described above, it does predict later grades moderately well a fair proportion of the time if the later grades are not too much later. But even this has a questionable circularity about it, showing only that similar kinds of judges will arrive at somewhat similar kinds of judgments about academic performance. The validity of both sets of judgments ought to rest on a different kind of evaluation.

Academic achievement tests appear to provide that different kind of judgment on which the validity of grades can be based. Scores on the Graduate Record Examinations, for example, sometimes predict graduate school grades moderately well and can themselves be predicted from the undergraduate gradepoint average (Lannholm, 1968b). But the circularity is not really broken because both aptitude and achievement examinations are constructed explicitly to predict grades and derive their validity entirely from them.

Westland (1969) concurs in stating that in order for college degrees, and by inference grades, to have social significance, they must have their meaning validated by social, not academic criteria.

I contend that at the moment we just don't know, in the scientific sense, what we are assessing. The problem is the criterion problem. We risk chaos if we don't look beyond our own ravel for justification of what we are doing (Westland, 1969, p. 360).

Woodring made a similar point, though not quite as cogently.

But no one can seriously believe that grades are the goal of higher education. And the assumption that those who make high grades are the ones who profit most from their education and are most likely to make the greatest contribution to society after graduation should be re-examined, for it must withstand a considerable amount of contradictory evidence (Woodring, 1968, p. 42).

Nevertheless, grades and gradepoint averages cannot be dismissed. The pooled judgments of intelligent people are a far sounder base for decision than is available otherwise. While Woodring (1963) contends that "grades have little meaning except as evidence of readiness for more formal education," Westland's view seems sounder. Grades probably do represent something useful; we just don't know what it is.

Selection for employment

The extent of the use of grades as selection criteria by employers is uncertain. Some put heavy weight on grades; others use them only for very coarse screening; still others use them not at all (Calhoun and Reddy, 1968; Committee on Grading, 1970; Dickenton, 1955; Kappel, 1962; Midwest College Placement Association, 1964-65; Paquette, 1966; Walters and Bray, 1963). Those employers who depend heavily on grades tend to have strong convictions but little evidence of their value. Law firm representatives interviewed by the Committee on Grading at the University of California School of Law, for example, were quite vehemently in favor of a detailed grading system, stating that students much below the top of the class just would not be adequate for work in their firms (Committee on Grading, 1970). Yet their failure to hire any but top

students makes one wonder on what evidence that policy was based.

Hoyt (1965, 1966, 1968) reviewed the studies he could find up to 1965 that related college grades to occupational success. The studies were scarce and their results equivocal. Hoyt concluded that "college grades have no more than a very modest correlation with adult success no matter how defined (Hoyt, 1965, p. 45)." Studies reported since Hoyt's review provide evidence on both sides of the issue but his overall conclusion remains valid (Calhoun and Reddy, 1968; Heckman, Banas, Lazenby, and Moore, 1969; Kinloch, 1969; Mason, 1965; McClaine, 1968; Pigge, 1968; Porter, 1969; Salyer, 1968-69).

The low relationship between college grades and occupational performance does not mean grades are useless determinants for employment selection (Hoyt, 1966; Raimi, 1967). College grades should indicate the level of student performance in academic tasks associated with understanding a given body of knowledge. This kind of performance may or may not be similar to the performance required on-the-job. If the gap some businessmen see between the academic world and the practical world really exists, employers should not expect grades to be related to job success. In Raimi's view, job success depends much more on experiences and capabilities developed after being hired than on the few years of college courses that precede employment. Therefore, job experiences in a few years heavily outweigh any college effects (Raimi, 1967).

If it is true that the effects of college are soon overshadowed by employment experience, one might ask why some employers stress grades so heavily. One reason is that good grades may indicate a facility for learning that will help a person acquire the knowledge and skills necessary for good job performance. Another is that some of the knowledge and understanding acquired in college may be necessary as a starting point for developing the additional knowledge and skills required on the job. The relationship between college and job performance would then become attenuated with time, a phenomenon that has been observed (Kinloch, 1969). These reasons for the declining relationship, however, are suppositions neither supported nor refuted by evidence.

A possible reason that some studies show moderate and others negligible correlations between grades and job performance is the greater importance in some job settings of compliance or willingness to follow instructions uncritically. Some evidence exists that this personal quality is associated with grades (e.g., Domino, 1968; Holland, 1960; Pemberton, 1969) and its importance in some kinds of jobs may be presumed. The nonacademic qualities of agreeableness, personableness, compliance, and sensitivity to the instructor's preferences that please faculty members can also be expected to please job supervisors.

Other differences in job requirements may also account for the varied results in predicting job performance from college grades. Heckman, Banas, Lazenby, and Moore (1969) found correlations between grades and the salary

progress of managers in a large manufacturing company to be highest for engineers and lowest for those in purchasing and traffic departments. If, as seems likely, both grades and job performance are multidimensional, correlations between them will fluctuate widely, depending on how the determinants of each complex variable happen to be combined into a single measure, and what relationship exists between their primary components. If academic grades are used in employment selection, more needs to be known of the structure of both grades and job performance. Determining relationships among selected components of the two kinds of performance may be useful.

Motivating students

A second widely asserted purpose of grades is to act as "motivators"—that is, to induce students to apply themselves to learning things they would not learn if not graded. Students and faculty alike believe that grading does have that effect (Katz and Associates, 1968; Sparks, 1969; Stallings and Leslie, 1970), and studies of Pass-Fail grading have indicated that the nature of the grade does influence how students will allocate their study time (Erickson, 1967; Feldmesser, 1969; Freeman, 1969; Karlins, 1969; Milton, 1967; Morishima and Micek, 1970). But the available evidence is too superficial for conclusions about motivating effects of grades to be held with any confidence.

The studies cited above showing the effects of Pass-Fail grading on allocation of study time demonstrate that students put less effort into Pass-Fail courses than into other courses. Each of these studies, though, was concerned with optional Pass-Fail grading. The students were permitted to take one Pass-Fail course per term; all other courses were graded A through F. Almost invariably in these circumstances students slighted the Pass-Fail course. But this can hardly be considered a damaging criticism of Pass-Fail grading.

Pass-Fail options typically exclude courses in the student's major field. That students should emphasize courses in their major field at the expense of other courses, often taken only to satisfy an institutional requirement for breadth, should not be cause for concern. The opportunity given students to allocate their study time selectively seems as much an argument in favor of Pass-Fail grading as against it.

Evidence from studies of limited Pass-Fail options is inadequate to evaluate the effects of Pass-Fail grading applied throughout an institution. Where complete Pass-Fail grading or purely descriptive grading has been instituted, no evidence has been found that students put less effort into their studies than they would under any other grading system. Sarah Lawrence College has operated without grades for many years (Murphy and Raushenbush, 1960), as have a number of other liberal arts colleges. One department of the University of California Medical School was successful with a system of faculty comments instead of

grades for a number of years until the faculty, over strong student objections, returned to a more conventional system of grading (Marshall, 1968). At the University of Kansas Medical School a shift to Pass-Fail grading seemed to reduce competition between students to a slight extent but had no discernible effect on student effort. The contest of students pitted against faculty, in which the students work to get past the obstacles the faculty throw in front of them, continued. The contest between student and student to see who could outperform the other had never been great (Becker, Geer, Hughes, and Strauss, 1961). Horowitz (1964), also at a medical school, found no decline in student effort after all grades were abolished, but did find that the appearance of indolence or of lack of interest can both be misleading. Finally, at the University of California at Santa Cruz, where complete Pass-Fail grading has been the practice since the opening of the institution in 1965, faculty members saw no evidence that students worked less diligently than had students at other institutions having more conventional grading systems (Committee on Educational Policy, 1970).

The experimental program that comes closest to providing a useful comparison between a graded and an ungraded instructional system is that followed at six liberal arts colleges, in which selected students pursued a 4-year program of independent study without specified course requirements and without grades (Cole, 1966; Operation Opportunity, 1970; A Report on the Independent Study Program, 1970). Within the same institution, some students worked under the usual grading system while others were freed completely from grading requirements. This does not mean the students in the experimental programs were not evaluated; they were. But the results of those evaluations were communicated directly to the student without recording a grade. The consequence of these programs cannot be attributed to the absence of grades for two reasons. The students were carefully selected and many elements of the experimental program other than the absence of grades could have been responsible for its effects. Nevertheless, some inferences about the effects of grades can be drawn with no more recklessness than is involved in most of the current opinions about grades.

Some evidence of the effect of grades as motivators may be observed, in that students in the experimental program often chose not to do some of the things that would have been required in regular courses. They tended, for example, to do less writing than was required of other students. But they did study and they did learn, although probably in ways not as obviously well-ordered as some faculty members would have liked. At the end of the first 4 years of the experiment the graduates included a number of Phi Beta Kappas and Woodrow Wilson fellows.

A tentative conclusion from reports of the programs is that grades played only a small part, if any, in inducing students to learn. On the other hand, the examination procedures, whether a grade was to be assigned or not, did

guide the students' academic behavior. Impending examinations often induced intense anxiety, even though no grade was to be given.

The primary source of student discomfort in the program, which was often great, seemed to be neither the absence of formal grades nor even the lack of structure. Instead, it was the ambiguity of many aspects of the program, due partly to its newness. The students often were not sure what was expected of them, were not ready to believe that they could, with their perceptor's guidance, set their own expectations, and were uneasy over their own evaluation of their progress.

It is not surprising that abandonment of customary guidelines and indicators should lead to anxiety and discomfort. An unusual kind of student is needed to manage it. When an entire college, such as Sarah Lawrence or the University of California at Santa Cruz, changes the guidelines, the effects are much less severe. But even where the students were a very small group in a new and sharply divergent program, the absence of formal grades did not lead students to squeeze through with as little effort as possible. The students either performed well or voluntarily withdrew to return to a more familiar academic environment and to whatever constraints course grades impose. The program was clearly not an invitation to indolence.

One conclusion that seems justified is that different kinds of students respond differently to different pedagogical procedures. While some students need the formal affirmation of accomplishment that a final grade gives them and will direct their efforts toward that goal, others find the constraints of grades onerous. This should hardly be surprising and has been reported before (Becker, Geer, Hughes, and Strauss, 1961; Birney, 1964; Horowitz, 1964; Miller, 1967).

In several studies, students have been observed closely enough and over a long enough period of time for informed judgments about motivational processes to be made (Becker, Geer, and Hughes, 1968; Becker, Geer, Hughes, and Strauss, 1961; Horowitz, 1964; Murphy and Raushenbush, 1960). As in the experimental programs described above, students were intensely concerned about their academic performance as a basis for their own self-evaluation and the satisfaction that results from a sense of competence. But the information they used for self-evaluation came from a wide variety of sources, not just from grades.

Students' needs for formal certification of achievement are an externally imposed incentive to study. The desire to perform well simply for the resulting sense of satisfaction is more internally based. Studies of the Pass-Fail option suggest that the external reward may override the internal one. Yet the desire for competence, as assessed by the student himself and as revealed in a variety of ways by teachers and by other students, provides a strong motivational force in many students.

The distinction between extrinsic and intrinsic sources of reward has been given as one reason for the inadequacy

of grades as motivators (Committee on Educational Policy, 1970; Karlins, 1969; Miller, 1967). The extrinsic-intrinsic distinction, however, is not always clear. The student's own self-assessment and intrinsic satisfaction, as Becker, Geer, and Hughes (1968) and Horowitz (1964) have shown, depends largely on external sources. When self-evaluative procedures and opportunities are limited, as sometimes happens concurrently with a de-emphasis of grading, many students become uncomfortable. But the anxiety is likely to arise not from the absence of grades but from lack of an opportunity for self-assessment. Grades at the end of a course only act as formal confirmation of the self-assessments students have been making regularly. Disputes between students and faculty members over grades occur when the grade does not confirm the student's previously formed self-assessment.

So far the motivating effect of grades as rewards for which students work has been considered. But grades are also used punitively by faculty members to coerce students into class attendance, performance of assigned work, and general deportment of the sort that pleases the teacher (Buchman, 1970; Dressel and Nelson, 1961; Goodman, 1964; Mayhew, 1969; Schwab, 1954; Wallace, 1966). The reluctance of some faculty members to change the grading system seems due to a fear that without the coercive effect of grades the teacher would lose most of his influence over student performance (Mayhew, 1969). The possibility that students would not attend a professor's lectures or follow his directions for study if they were freed from the demands of grades can be a frightening prospect. Holding to grades to avoid facing that prospect is more comfortable.

In summary, the motivating effect of grades is complex and not well understood. Some students value the formal affirmation of accomplishment that grades represent and work to get it. For others the almost continual self-assessment derived from cues provided by teachers, other students and regular course activities is sufficient. Published grades at the end of a course have little additional motivational effect for these students.

Another point basic to the use of grades as motivators should be mentioned, although it will not be developed at length. It is the question of whether faculty members should be concerned at all with devices to induce students to study. As colleges increasingly abandon the role of surrogate parent with respect to the social behavior of students, coercing students into desired patterns of activity by faculty-administered rewards and punishments might also be abandoned as unnecessarily paternalistic.

The informative function of grades

The first two purposes of grading—as selection devices and as motivators—can both be considered services primarily to society rather than to students. The use of grades as selection devices permits higher education to perform its function as a social sieve, determining who shall be

admitted to positions of prestige, power, and financial reward (Caplow, 1954; Clark, 1962; Jencks and Riesman, 1968; Mayhew, 1969; Sexton, 1967; Sparks, 1969; Tyler, 1969). Grades as motivators also serve society's purposes, inducing students to kinds and levels of performance they presumably would not choose freely. Although some students benefit from the use of grades in selection, and although grade-induced studying may also be considered beneficial to the student, the primary service is to society. In contrast, the third most commonly discussed purpose of grades, which receives far less attention than the first two, is their use as a device to serve students by informing them about their performance.

The contention that feedback to students about their performance constitutes an important purpose of grades (Committee on Grading, 1970; Dale, 1969; Sparks, 1969) confuses evaluation (the assessment of performance) with grading (reporting of the assessment results). A variety of procedures are available to inform students about the nature of their performance without the publication of a summarizing symbol to represent overall performance in a course.

Becker, Geer, and Hughes (1968) and Horowitz (1964) found that students used a variety of cues to assess their level of performance relative to other students. Stallings and Leslie (1970) reported a survey of students at the University of Illinois in which most students did not consider grades to perform a useful feedback function. Students are seriously concerned with self-evaluation and tend to become anxious in the absence of evaluative information about their performance (Funkenstein, 1968; Horowitz, 1964). But course grades, since they do not appear until course completion and are limited in content to the information that can be carried by a single symbol, are not effective feedback devices. The evaluative procedures that lead to the most effective feedback are often not those that lead to the most useful ranking of students (Bloom, 1968; Husek, 1969). And relative rank on a global evaluation is not very informative at best.

Effective feedback helps students judge their progress on their own terms. Acquiring a general grasp of the major issues may be all a student wants from a course outside his major field but is far from adequate in a course important to his major. Effective feedback also leads to modifications in student behavior that will improve performance or to assurance that performance is adequate. It should indicate areas of weakness or topics insufficiently understood. Successful and unsuccessful methods of study should be identified soon enough to permit adjustments to be made.

Feedback should be related to the processes as well as the products of learning, differentiating among various forms and areas of academic accomplishment and indicating directions for future study. It is most effective when considered in relation to the student's previous accomplishment and capabilities. Performance in relation to other students has limited usefulness for feedback and is at times misleading, as when the other students in a class do not

constitute a useful reference group for some particular student. In Scriven's terms, feedback is a product of formative evaluation, grades of summative evaluation (Scriven, 1967).

If the objective of evaluation is to rank students for some purpose that requires a relative assessment of overall accomplishment, observation of the procedures a person goes through in arriving at a result is not important; whether he arrives at the desired or correct result is important. Final course grades, constituting a coarse ranking of students, indicate roughly what a student has accomplished academically in that particular course compared with other students in the course. They convey useful information primarily to people who were not engaged in the course. Students and teachers learn little from them.

For these reasons, conveying information to students should not be considered an important function of grading. The information grades convey to students tells them what information admissions officers and employers will have about them on which selection decisions may be based. Grades also convey the instructor's overall judgment of the student's total performance, which may help him decide about future work in the field of the course. But this is a low order of information in comparison with what the student has learned of his capabilities directly throughout the period of the course. The educational function of grades is therefore limited, both because they are assigned after the learning is completed and because they are little more than general summaries of information students have probably already received by other means.

Institutional purposes of grades

One of the major administrative purposes of grades is in selection to graduate and professional education, a purpose that does not directly serve the interests of the institution awarding the grades. The grade-awarding institution does use grades, however, for a variety of internal administrative purposes. The most important is probably in decisions about whether to permit students to re-enroll in succeeding terms. Although this use of grades is critical for only a small proportion of students, over a period of years it excludes large numbers of students from further education. Its total social effect is therefore substantial, constituting an important way that the segment of the population permitted access to higher occupational, economic, and social positions is defined.

At three stages in the educational process—admission to college, retention in college, and admission to advanced education—grades exercise a substantial influence on decisions about who shall be permitted to continue. The assumption that grades constitute a defensible basis for these decisions has some rational justification. Teachers prefer students whom teachers before them have preferred. But whether the elements of performance that determine teacher preferences coincide substantially with the elements

of performance on which decisions about continued education should be based is a question that has not been examined. As Scriven (1969) and Stake (1970) have urged, research is needed to determine how decisions about allocation of limited educational resources can most justifiably be made.

Other institutional purposes are in determining admission to advanced courses, eligibility for extracurricular activities, awarding financial aid, and awarding academic honors. In these areas grades may be a sound basis for decision. The awarding of academic honors, for example, is by definition based on grades. The award of scholarships and other financial aid on the basis of grades is more questionable. Typically, a student who needs financial assistance has the form of that assistance—whether outright grants or loans and part-time work—based on his grades. The justification for this practice is similar to but less defensible than the justification for basing selection on previous grades. The rationale—that better performing students, in terms of the behavior indicated by grades, are more deserving of financial help than other students—cannot have been adequately examined in view of the limited knowledge of what grades represent.

Grades are said to provide important information to teachers, permitting them to judge their own effectiveness, and to department heads and other administrators, permitting them to make comparative evaluations of teachers and departments. Grades are completely unnecessary, however, to teachers' self-evaluations. Evaluation of student performance is essential; grading is not.

Similarly, teachers, departments, and divisions are often compared with respect to their grade distributions. These comparisons, however, provide no more information than how teachers, departments, and divisions compare in the grade distributions they produce. How this information is to be interpreted is largely unknown. Whether consistently low grades in a department result from poor students, poor teaching, an inappropriate combination of teaching method and student characteristics, poor evaluation, or inordinately high standards cannot be determined from comparisons of grade distributions.

While the intended purposes of grades have seldom been discussed except by implication, the unintended side effects of grades have frequently been reviewed in detail (Becker, Geer, and Hughes, 1968; Marshall, 1968; Miller, 1967; Milton, 1966; Milton, 1968; Raimi, 1967; Trow, 1968). A large body of empirical data could be brought to bear on the intended but unexamined purposes. Very little of the extensive discussion about the unintended effects of grades is based on systematic observations.

This is not to say that comparing grade distributions is useless. It may suggest why student attrition is so great in one department. Unusually low grades assigned consistently by the same teacher may indicate a particularly critical, demanding instructor or they may suggest an underlying attitude of hostility toward students that interferes with instruction and learning. Additional information might then be gathered to determine the reasons for unusual grade distributions. But grade distributions in themselves say almost nothing about the teaching or learning that occurred.

Grading as preparation for life

Grades have occasionally been said to be desirable in preparing students to face the competition they will inevitably meet in the "real world" beyond school. This view seems to be a relic from an earlier day in which college was a pleasant, undemanding way for sons of the social and economic elite to spend a few years before moving fully into the adult world. Whether or not it was ever widely justified, it certainly is not today. To consider college experiences as not belonging to the "real world," whatever that may be, seems absurd.

Few nonschool situations, in employment or elsewhere, have anything resembling the grading procedures of college. Even in employment, evaluation through the use of written tests is not particularly common. Civil Service procedures may come close to some aspects of college grading, but the Civil Service is not typical of most employment situations and its similarity to college is limited.

A vast amount of evaluation does go on in almost every kind of situation, but most of it is highly informal, ad hoc, and far removed from anything like college grading. Yet even if situations were common outside college in which grading much like that in college occurred, this would not in itself give colleges the responsibility to prepare students for those situations. Even colleges that assert one of their purposes to be preparing students for life do not claim to prepare students for every kind of situation they may face. Preparation for the competition of examinations and grading does not have demonstrable value.

V. UNINTENDED EFFECTS OF GRADES

In illustration, a large amount of information is available that bears on the use of grades in selection to more advanced educational programs. Yet the philosophical, social, educational, and economic justifications for the use of grades in selection which could be examined in the light of that information, have been almost ignored. In contrast, little more than personal impressions, at times probably well-founded but at other times not, can be drawn on in support of the widely discussed view that grades distort the

learning process. The known is ignored; the unknown is described in detail.

Distortion of learning

Distortion of the learning process can have a variety of meanings. One is the belief of a large number of students that the kinds of activities that produce good grades are often not those that would produce optimal learning (Becker, Geer, and Hughes, 1968; Education at Berkeley, 1966; Katz and Associates, 1968; Miller, 1967). The need to maintain a grade-point average high enough to assure selection to graduate school, or permission to re-enroll for the next academic year, is presumed to demand student time and attention that could be spent more productively. Specification by the instructor, either explicitly or implicitly, of the details of what must be done to pass tests, write acceptable papers, take part satisfactorily in class discussions, or otherwise perform in ways that will be rewarded with good grades constrains student behavior to uniform tasks that may not be uniformly effective for all students (Cole, 1966; Miller, 1967; Milton, 1967; Torbert and Hackman, 1969).

The above argument does not contend that teachers should not direct the learning activities of their students. It does contend that students are capable of greater discretion than is allowed by the present system of grades in the ways they will respond to direction from the teacher. When the fact of college graduation was in itself the critical determinant of entry to desirable jobs and higher social status students were freer to control their own academic behavior. The "gentleman's C" was often an acceptable level of performance, and an occasional D was no more than a temporary blow to self-esteem. With the mounting importance of graduate education and of the grade-point average as a ticket of entry, student discretion in their academic activities has been severely curtailed.

The increasingly common flirtation of colleges with Pass-Fail grading has its origin primarily in the desire to give students wider latitude in their selection of courses. The reason most often given for introducing Pass-Fail options has been to free students from the constraints imposed by fears that courses in unfamiliar areas might damage their grade-point averages. Ironically, the most common objection to the consequences of introducing a Pass-Fail option has been to students exercising independence in another way—in their allocation of study time and effort.

Prescribing in detail what students must do to earn a satisfactory grade takes from them the responsibility for deciding what is important. The importance of the grade-point average, which gives force to the instructor's prescriptions for learning, prevents students from experimenting, exploring different approaches, and learning that some approaches will not work. But students are also prevented from learning that some approaches other than the instructor's may work admirably for them. The present

grading system therefore inhibits learning by not permitting failure, or by making failure too costly for students to experience (Torbert and Hackman, 1969).

A different kind of distorted learning results when students behave in ways unrelated to substantive learning, in forms of behavior calculated only to please the instructor. Asking the right kind of questions, feigning interest in the instructor's favorite topic, learning the style of answer the instructor prefers, and other purely grade-oriented ploys may not be totally useless with respect to substantive learning, but their intrinsic value is limited. The grading system is said to be the primary cause of dissipation of student effort in this kind of sterile, game-playing activity (Axelrod, 1968; Becker, Geer, and Hughes, 1968; Lavin, 1965; Raimi, 1967; Torbert and Hackman, 1969).

Bloom (1968) has suggested a third kind of constraint: grades may impose on learning. When teacher and students alike start a course expecting that only a few will learn enough to earn a top grade, and that some will learn no more than enough to get a marginal grade or worse, the expectations become self-fulfilling and reduce the aspirations and performances of both teachers and students. A prior history of earning average grades may put a ceiling on student expectations and performance. Bloom contends that most students in any particular class are capable of achieving the goals of that class. The most effective procedure and the time required for mastery of a course may vary, but instructional procedures should be capable of providing for variability in student predilections.

The present grading structure, in requiring that a learning period is to end and grades are to be assigned after a fixed period of time, imposes another constraint on learning (Fordyce and Bromley, 1969-70; Raimi, 1967). Successive courses in the same field are intended to be integrated, the first leading into the second and the second building on the first. When this occurs, the arbitrary ending of a period of learning after a fixed number of weeks may not be serious. But the adequacy of the integration of learning that has been structurally fragmented has been questioned (Sparks, 1969). The requirement that a student, after a fixed period, either move on to the next learning episode or repeat the entire process he's just been through seems dubious. Requiring students who have mastered a course in less than the allotted time to continue to go through the exercises of that course instead of moving on seems equally questionable.

The grading system is not the only reason for organizing learning into fixed periods of time. Some limitations are necessary simply because of the need for one teacher to accommodate a number of students. But present uses of the grading system in the selection and classification of students require that grades at least have the appearance of quantitative as well as qualitative comparability. If grades of students who took different courses in the same subject are to be compared, the two courses must have some kind of equivalence. Standardizing the time spent in the two courses provides that equivalence. The retention of grades

as selection devices therefore acts, with other considerations to inhibit the introduction of greater flexibility into the structure of education.

The lack of substantive meaning in grades (Erickson, 1966) is the primary reason for attempts to keep them, in some sense, equivalent. Because grades have no content other than the name of the course to which they are attached, grades would have virtually no meaning without the comparability, limited as it is, provided by the number of weeks of instruction that a grade represents. A grade does not indicate what a student knows, for example, of the effect of a regulated economy on competitive equilibrium, but only that he completed a course in economics somewhat more (or less) satisfactorily than most students in the course. Since they cannot be compared with respect to the substantive learning they represent, grades in two economics courses can be made comparable only in terms of the amount of classtime spent in each course.

If reports of student performance were descriptive, with respect to the substance and the level of performance, strained attempts at equivalence would be unnecessary. Persons using grades in decisions about selection would still be faced with developing some index of overall performance or suitability from descriptive reports that would often not be comparable. But this would not be an added burden. It would only represent a shift of that burden from those who teach to those who select. And those who select would be able to specify their own criteria instead of assuming that those used by the teachers were appropriate. The lack of comparability that would appear in descriptive reports is fully present in current grades and gradepoint averages; it is only hidden by the failure of grades to convey any substantive meaning. When no meaning is conveyed variation in meaning cannot be observed.

Becker, Geer, and Hughes (1968) described grades as "the major institutionalized reward available for academic work." In their view, grades act in college the way money does in society at large, as a medium of exchange for both tangible and intangible valuables, but primarily intangible ones in the case of grades. Grades therefore constitute a major element of the social environment to which students must accommodate. Their influence is ramified through most aspects of student behavior, beyond classroom and study activities into such areas as dating behavior and informal relationships between students.

The complete faculty control of the exchange of grades for academic performance puts students in a position of subjection. Thus one of the commonly stated goals of liberal education—training students to be intellectually self-directing—is subverted. Yet students retain some autonomy and can, through collective action, resist faculty demands with some effectiveness (Becker, Geer, and Hughes, 1968).

Wallace (1966) reported changes in attitudes toward grades over the period of an academic year that emphasize the role of collective action by students and the socialization of students through interaction with other students

Faculty members and nonfreshman students differed substantially with respect to the importance attached to different orientations toward college. The faculty members valued grade-oriented activity more highly than did the students. Over the course of the freshman year, the grade orientation of freshmen moved away from that of the faculty to a position consistent with that of the non-freshman students. Conflict between student and faculty expectations with respect to grades and the power of the socializing effect of the students are both indicated.

Distortion of teaching

The need for some equivalence in grades prevents instructors from varying their course content too far from a generally accepted standard for that type of course (Miller, 1967). Many professors believe an A in a course cannot be given without mastery of certain areas of content that would be agreed on by other professors in the field. Whether a particular professor agrees with the presumed consensus among his colleagues or not, he may feel his reputation endangered if he sends out students with A's who could then be discovered not to have mastered some content area. The argument can be made that grades, in placing this kind of constraint on professors, are a desirable device for maintaining academic standards. But Miller (1967) lists this effect of grades among their deficiencies. Whichever view is taken, the function of grades in imposing instructional constraints is important enough to be examined. It has not been.

The requirement that grades be given, in a certain form and representing certain presumed accomplishments or capabilities, is considered a major cause of a sterile but common conception of teaching (Axelrod, 1968). An instructor faced with a requirement to order his students at the end of a fixed number of weeks with respect to their relative accomplishment in his course is inclined to organize the course in such a way that grading can be accomplished simply and can be defended against attack by the students. This often leads to common requirements for all students, the setting of tasks that can be carried out mechanically and therefore easily observed, and authoritarian control of the activities of the students.

When instructors are required to assign ratings of merit to students that will affect later decisions about those students, they are put in the role of judge rather than mentor. If the two roles are incompatible, as has been contended (Axelrod, 1968; Marshall, 1968; Mayhew, 1969; Raini, 1967), then current grading practices must interfere to some extent with learning. Students have great difficulty ignoring the fact that their teachers will at some point grade them. They leave questions unasked rather than risk displaying ignorance. They stifle critical comments that might lead to profitable clashes of ideas. They stay within the instructor's guidelines instead of stepping outside them when an approach that looks intriguing has either already been rejected by the instructor or has not occurred to him.

A large number of sometimes subtle but important differences can be found between the behavior of someone being taught and someone being judged.

In an analysis of the development of two new experimental colleges, Riesman, Gusfield, and Gamson (1970) described the effect of grading on faculty behavior and relationships much as Becker, Geer, and Hughes (1968) had done with grades and student behavior. "Grades serve not only to sort and certify students but, more symbolically, to sort and certify faculty vis-a-vis one another (Riesman, Gusfield, and Gamson, 1970, p. 137)." The nature of the student-faculty relationship, of the responsibility of faculty for students, was reflected in the grading behavior of the faculty. Interfield conflicts with respect to grading philosophies developed in which students were able to play one faculty point of view against another. Some faculty members were put on the defensive, which ones depending on the prevailing attitude toward grading and on the goals of the institution as perceived by the faculty at large. The role of grades as an implicit affirmation of faculty values gives them an importance in faculty relationships not often acknowledged.

Student attitudes and behavior

One of the common complaints about grades is that they produce unnecessary anxiety in students (Benson, 1969; Committee on Educational Policy, 1970; Funkenstein, 1968; Karlins, 1969; Pass-Fail Study Committee 1969; Raimi, 1967). Whether anxiety is desirable or undesirable in a learning situation is a complex question. Personal attributes of the student, the nature of the learning task, its importance to the student, and the level of anxiety induced all interact to produce widely varying effects. The only statements to be made with reasonable confidence about grades and anxiety are that the anticipation of being graded does raise students' anxiety levels and that anxiety is usually unpleasant. These two facts probably account for students' overwhelming endorsement of Pass-Fail grading in preference to conventional grades.

The introduction of a competitive atmosphere to campuses and classrooms is attributed to grades (Becker, Geer, Hughes, and Strauss, 1961; Bloom, 1968; Karlins, 1969; Miller, 1967). Its effects are considered both desirable and undesirable and, like those of anxiety, are probably mixed. Those who consider competition desirable say it provides a valuable motivating force and gives students useful experience in handling competitive situations. Others say it interferes with learning by inhibiting student cooperation and collaboration, by adversely affecting students' peer relationships, and by lowering student morale.

Cheating is said to be a consequence of grades (Birney, 1964; Raimi, 1957) and may be one reflection of an atmosphere of competition. One of the contentions of proponents of Pass-Fail grading is that cheating is less

prevalent with that system than with conventional grades (Committee on Educational Policy, 1970; Stallings and Leslie, 1970). So far as is known, however, systematic observations of the relationship between grading and cheating have not been made.

Students' decisions about graduate study were hypothesized by James A. Davis (1966) to be affected by their grades as undergraduates, relatively low grades acting to discourage students from applying to graduate school. According to Davis's theory, the selectivity of the undergraduate college would not be given much consideration by the students. Average students at selective colleges would then give up graduate school aspirations even though they may be superior to top students at mediocre schools who had their graduate school aspirations strengthened by an undergraduate performance that was high only in relation to a mediocre standard. Davis presented evidence from a large-scale survey of college graduates that partially supported his view. Werts and Watley (1968, 1969) provided some confirmatory evidence for the theory, although Werts (1968) raised questions about the adequacy of the analysis for the purpose. Davis pointed out that the effect of the process, if it occurs, would be to ensure the presence of capable people--the mediocre students at the excellent colleges--in occupations of relatively low prestige, such as teaching.

Social effects of grades

A largely unexamined but highly important aspect of grades is their effect on the social structure. The view that grades are a mechanism by which education maintains the existing class structure, controlling access to higher social and economic levels, has been discussed earlier. Major proponents of this view are Caplow (1954), Katz (1968), Jencks and Riesman (1968), and Sexton (1967), but others who have raised questions about the socially conservative effects of grades are Erickson (1967), Hoyt (1966), Lavin (1965), and Tyler (1969). Clark (1962) considered the socially constraining effects of education to have been reversed in the present century as education became more widely available. He developed the widely held position that education acts as a mechanism for upward social mobility and for reordering social positions in accordance with individual merit rather than social origin. Whichever view of the social effects of education is more accurate, grades are an important mechanism for producing those effects.

Few of the possible effects of the grading system are as important as its role in either maintaining or reordering social and economic positions. This alone should justify far more intensive study of the grading process than has been carried out. Most of the evidence on the effects of grading consists of student reports of feelings or attitudes. Students say they feel anxious about grades; but the level, effects, and precise source of the anxiety are unknown. Some

students and faculty members say that grades interfere with learning, supporting their statement with plausible arguments but few pieces of evidence. That an educational

practice as important, as pervasive, and as much the subject of contradictory views as grading should have had so little systematic investigation is startling.

VI. TECHNICAL ISSUES IN GRADING

Most of the preceding discussion of the forms, purposes, and effects of grades has been concerned with issues external to grades themselves. Yet the intrinsic characteristics of grades—the processes through which academic performance is judged, the ways those judgments are translated into scaled symbols, and the composition and stability of both the judgments and their translations into symbols—to a large extent determine how well grades perform their external functions.

Multiple components versus a single dimension of performance

Academic performance can be considered the result of some amalgam of inherent intellectual capability, possession of relevant information, intellectual curiosity, perceptiveness, analytical power, ability to synthesize concepts into higher order abstractions, clarity of exposition and expression, and other intellectual capabilities. Attitudes and behavioral tendencies add more elements to academic performance. Industriousness, commitment to an academic field, responsiveness to instruction, intellectual integrity, and some other attributes of personality are difficult to distinguish from variables that are more explicitly academic. Finally, most professors respond favorably to some student attributes, such as physical attractiveness, pleasantness of manner, or apparent earnestness, that are irrelevant to academic performance but that sometimes color judgments of performance. What peculiar combination of these and other variables is reflected in an instructor's evaluation of student academic performance is never completely clear, even when a course grade is determined entirely by the mechanical accumulation of points on a set of examinations. Different kinds of student performance reflected in tests given at different points in a course, for example, might add to identical totals and identical grades for two students who differed sharply in the nature of their performances. Although different kinds of performance may be equivalent with respect to overall level, that determination is seldom made and its implications seldom explored. How much expository fluency is equivalent to how much analytical skill is the kind of question too lightly passed over in determining grades.

The multifaceted nature of academic performance has been offered frequently as a major problem in the interpretation of grades (e.g., Ebel, 1965; Milton, 1966; Milton, 1968; Trow, 1968). Evidence that grades are

determined by various kinds of behavior is not hard to find. Faculty members vary in the weight they give to such aspects of performance as effort and improvement (Axelrod, 1964). Medical school grades have been shown to be multidimensional (Haley and Lerner, 1967). English professors differ in the qualities they observe in assigning grades (Lewis and Smith, 1969). Faculty members in the physical sciences differ from those in the social sciences in the expectations they hold for students (Gamson, 1967; Riesman, Gusfield, and Gamson, 1970). The diversity of academic performance seems incontestable.

When a set of grades, each determined by a somewhat different set of attributes, is averaged, the qualities represented by that average can only be guessed at. The argument has been made that averaging course grades is desirable because it compensates for the variable nature and uncertain assessment of the student attributes, capabilities, and performances that determine individual grades (Bramer, 1970; Dale, 1969). Deviations from the average of the judgments of 20 to 40 instructors are said to cancel themselves out, leaving a reasonably stable indicator of whatever is common to most faculty evaluations. But the nature of that common core is hard to identify.

Boldt (1970) has recently provided empirical support for the existence of a single dimension underlying performance in a number of courses. At two different graduate schools of business, variation in student performance across 31 and 70 different courses could be accounted for almost as well by one dimension of performance as by two or three. Even though some courses were quantitatively oriented while others were heavily verbal, performance in those two types of courses could not be clearly differentiated. Boldt concluded that "the present study uncovers no reason to reject gradepoint average as a simple and descriptive index of achievement (p. 23)."

In spite of the study's limitation to graduate courses in business, in which about 90 percent of the grades were either A or B, substantial support is given to the view that gradepoint averages represent quite well some composite of whatever kinds of academic performance are reflected by grades. But the nature of that composite dimension—the components of student performance that it combines into a single measure—remains undefined. Its usefulness beyond predicting second-year grades from first-year grades would be enhanced if its components and their interrelationships were known.

Another study related to grading in graduate schools of business suggests that academic achievement may be treated

either unidimensionally or multidimensionally, depending on the situation. A set of rating scales was developed that described 13 qualities faculty members in graduate schools considered desirable in their graduates. A total of 191 first-year students at two schools were rated on the 13 qualities by 27 different faculty members, with each faculty member rating up to 10 students (Hilton, Kendall, and Sprecher, 1970).

Five of the 13 attributes—perspective and breadth of knowledge, technical knowledge, critical awareness, problem analysis ability, and communication skill—can be considered modifiable by instruction. The other eight, such as persistence, initiative, and flexibility, seem less accessible to instructional change but may nevertheless affect judgments about student performance. These 13 desirable attributes, but particularly the five subject to change under instruction, might be expected to vary in importance across courses and appear as distinct dimensions in studies such as Boldt's. But the ratings of students on the five modifiable attributes were all highly interrelated and were all moderately related to first-semester grade point averages.

For some purposes, these qualities may be treated as distinct attributes, but they can also be considered as somewhat different components of a single dimension of academic performance. The choice between considering performance a single dimension or several should depend on its use. Present practice is to treat academic performance as a single, global entity. Greater knowledge about its components and their relationships to other kinds of performance should lead to better student evaluation, better grading, and more effective use of grades.

Reliability or consistency of grades

The reliability of grades can be observed in several ways, each involving some aspect of consistency. The internal consistency of grades is a measure of the degree to which the various observations made by a particular instructor to arrive at judgments about the grades of students in one of his courses reflect a common form of academic performance. For example, if the instructor's evaluations, whether of written papers, objective examinations, or observations of classroom performance, all depend heavily on the recall of factual material, his grades are likely to show a high index of internal consistency. Course grades can also be consistent—that is, show the same relative ordering of students—across instructors teaching the same course, across different courses taught by the same instructor, across different classes taught by the same instructor in the same course, and across time. Most of these situations, such as different instructors teaching the same course to the same student, exist only hypothetically, but they illustrate the varied meaning of reliability. Consistency or reliability in any of these other forms is limited by the internal consistency of the grades of individual instructors.

The reliability of grades is clearly related to their dimensionality. As the number of attributes considered in assigning grades increases, the reliability is likely to decrease. Reliability can stay moderately high, however, if the various attributes observed by an instructor are themselves highly related. The high reliability of grades across courses and instructors, for example, in spite of differences in course emphases and methods of evaluation, is probably due largely to the common element of verbal ability in most academic evaluations.

Two studies in recent years have measured the reliability of grades, using different procedures but with similar results. Clark (1964) defined reliability as the ratio of the variance of individual grade point averages to the total variance of all grades. For 18 classes of freshman women at Northwestern University from 1931 to 1959, with an average of almost 300 women in each class, reliability coefficients ranged from .70 to .80 with a median of .74. Barritt (1960, using a simpler but somewhat analogous computational procedure that consisted of computing correlation coefficients between random halves of students' grades, found freshman grades for 237 students at Indiana University to have a reliability of .84.

Both these studies show that a large part of the information in freshman grades can be associated with some unidimensional concept. They do not rule out the usefulness of a more complex conceptualization of grades. They do indicate, with Boldt (1970), that treating grades as though they represent a single, general kind of academic performance is a sound procedure.

The common assertions about the unreliability of grades usually refer to the inconsistency across faculty members with respect to their judgments about the quality of a particular piece of student work, usually a written paper or essay examination (e.g., Stewart-Tull, 1970). This source of inconsistency may be due either to variations across instructors in the attributes they consider important or to inability of faculty members to make consistent judgments.

The temporal stability of grades can be affected by inconsistency of any sort. In a study of grades at the University of Illinois between 1962 and 1966, correlations between grades in adjacent semesters were moderately high, around .55 to .60 (Humphreys, 1968). The correlations of first-semester grades with high school rank and an achievement test were both about .50. But the correlations of the same predictors with grades in each succeeding semester declined regularly and dramatically. High school rank showed a correlation of .22, for example, with grades in the eighth semester of college. Similarly, the correlations between grades in different semesters declined regularly from about .54 to .34 for a constant sample of students as the time between semesters increased. A virtually constant standard deviation through all eight semesters dispenses with the possibility that reduced variability in the later semesters could account for the declining correlations.

While grades are consistent across courses in any one semester, they are not very stable over an extended period

of time. Either the nature of the academic performance that grades reflect is modified as students progress through college, or else students fluctuate substantially from year to year in their performance. In either case, predictions of performance beyond the next academic year are dubious. Also in either case, knowledge of the nature and behavior of the components of academic performance would be valuable.

Adjusted grade distributions

Faculty members tend to be concerned about the problems of evaluating students and arriving at grades that will summarize performance on the variables considered important in a particular course. How a student performs in other courses and the nature of the performances considered important in other courses are clearly irrelevant. Variation across courses in the *nature* of student performance is accepted as inevitable and proper and attracts no one's concern. Variation across courses in the *level* of performance judged to have been reached by the students does attract attention and is considered undesirable (Juola, 1968).

The concern over variation across courses in the average level of student performance and the acceptance of variation across courses in the nature of student performance are not necessarily inconsistent. The desire for students in dissimilar courses to have similar distributions of grades is no more than a desire for the grade scales to be comparable in different divisions, departments, and courses. Only if that is the case can the grades of all students in all courses be considered equivalent and capable of being summarized in a gradepoint average.

The legitimacy of gradepoint averages, in contrast to individual course grades, is the concern of deans, of faculty members when they are serving on admission committees, and of directors of institutional research. That concern is reflected in the suggestion that the grade distributions within any class be adjusted to the capabilities of the students in that particular class, as indicated by academic aptitude test scores or previous grades or both (Anderhalter, 1962; Berdie, 1965; Fricke, 1965; Grant, 1956). A class composed predominantly of A and B students would receive a predominance of A's and B's. A class that included a wide range of capabilities would receive a wide distribution of grades. Grades would therefore be approximately equivalent across all classes.

Computers make it possible for each instructor to receive, soon after the start of a term, a report of the distributions of the previous grades and test scores of the students in each of his classes. He need not have that information about individual students and is under no obligation to assign any particular grade to any individual. But he would know the general level and range of performance to be expected in each class and could adjust the eventual grade distribution of each class accordingly.

As Fricke (1965) pointed out, adjusting grade distributions to student capabilities is an effective resolution of a problem of grading that is distinct from problems of evaluation. Judging the nature and quality of students' performances is evaluation and results in a ranking of students within the group evaluated. Then the determination of where to place those students on some grade scale—how many, if any, of the top students merit A's, how far down the scale the lowest student falls, and where the students in between belong—is a new problem: grading. Adjusting grades seems to assure comparability of the grade scale across departments and classes differing in student capability by anchoring the grades in any class to scores on a common aptitude test taken by most of the students or to the averages of the students' previous grades. But its independence of the evaluation process accounts for at least two deficiencies in the procedure, and others exist as well.

First, a common scale can be used in the assessment of dissimilar objects only if the objects possess some common attribute. No manipulation of numerical scales can make inherently different concepts equivalent in any very useful sense. The academic tasks in chemistry classes are different from those in literature classes. Adjusting chemistry grades and literature grades in accordance with a common anchoring variable is justifiable only to the extent that the anchoring variable is associated with both chemistry and literature. Distinctions between performance in chemistry and literature will be systematically de-emphasized, even though the areas that distinguish between the two fields may be those most worth emphasizing. Any other source of variation in the meaning of grades, as well as differences across fields, such as basing some grades on the recall of factual material and others on comprehension of complex relationships, further detracts from the comparative meaning of grades in a way that cannot be remedied by adjustments to the grade scale.

A second objection was raised by Gold (1966), who pointed out that academic performance is a consequence of the activities of teachers as well as students. Under Fricke's proposal, the relative achievement of two comparable classes, one brought to a high level of performance by an outstanding instructor and the other left at a relatively low level of performance by a poor instructor, would be indistinguishable. Appropriate evaluation would reveal the differences in performance—grading adjustments would hide them.

Two other objections can be made to adjusted grades. One is that the process is biased against recognition of change in performance. Even though an individual student's grade is free to take any level, the number of A's in a class is constrained by the previous performance of the class as a whole, and a sharp growth in general student interest and performance will not be recognized in the grades students receive. If grades have a motivating effect, as many contend, that effect might be curtailed or even reversed; the lack of recognition of improved performance could discourage further improvement. Although this potential

damping effect of grade adjustments on changes in performance might be negligible, no evidence is available on which to base such a judgment.

The final objection is that grade adjusting requires the assumption that criteria of performance are constant across courses and from one year to the next. If grades in a particular course are intended to reflect an understanding of complex interrelationships in the flow of history, for example, adjusting them to previous grades that indicate a variety of kinds of performance, including things like remembering taxonomies or applying rules of integration in calculus, makes little sense. Adjusting grade distributions to previous grade-point averages gives disproportionate weight to those few elements of academic performance, whatever they may be, that are common across all courses.

Interactive determinants of grades

Other writers, as well as Gold (1966), have pointed out the complex, interactive nature of the determinants of grades. Erickson (1966) described grades as the result of "an extremely complicated interaction between a teacher, students, and a body of knowledge." Haagen (1964) added the effects of the institutional climate and of society at large, but also stressed their interaction with student and instructor characteristics. Variations in faculty standards (Axelrod, 1964; Juola, 1968; Kirby, 1962; Trow, 1968; Webb, 1959), in departmental standards (Aiken, 1964; Anderhalter, 1962; Gamson, 1967; Juola, 1968; Kelly and Thompson, 1968; Pemberton, 1969; Trow, 1968), and in average student capabilities from year to year (Aiken, 1963; Bowers, 1967; Hills and Gladney, 1968; Miller, 1969; Webb, 1959) are all influences on grades that are beyond the student's direct control. If grading is to be free from effects not under the student's control, some approach to an absolute standard is necessary.

Absolute versus relative standards

Inequities in relative grading standards, due to any of the sources of variation beyond the student's control, may be avoided by establishing absolute standards and making each student's grade independent of any other student's grade. Although relative grading standards and grading "on the curve" have been dominant over absolute standards for half a century, a resurgence of interest in absolute grading is occurring in the guise of criterion-referenced testing (Ebel, 1962; Richards, 1970). Work in programmed learning requires the determination of absolute levels of performance to direct the learner to the next stage of instruction. Individualized instruction in any of its varied forms similarly requires absolute scales of performance.

One consequence of an absolute grading standard is the opportunity to avoid the fixed time period of a semester or quarter in evaluating achievement. If, as is generally

acknowledged, students learn at different rates, permitting students to use varying amounts of time to reach a desired level of achievement seems preferable to the present system of applying the same evaluative standards to all students at the end of a predetermined number of weeks (Bloom, 1968; Dressel and Nelson, 1961). Evaluation and certification of achievement at the end of variable periods of time require the development of absolute standards.

Yet the argument over relative or absolute grading standards is to some extent a false issue. Even in the British system of external examiners and in criterion-referenced testing, the "absolute" standard is established in relation to some expectation of performance based on past experience with examinees in similar circumstances. The real issue is in specifying the source of the standard on which grades are to be based. Neither a narrowly defined relative standard that results in a fixed distribution of grades throughout a class regardless of the general level of the class's performance, nor a rigid standard based on scores on a standardized, externally administered test, seems desirable. But the decision as to what standard should be applied must be reached with some care, and that decision cannot reasonably be reached without consideration of the purposes for which grades are to be used.

External versus internal evaluations

Consideration of absolute standards suggests placing the responsibility for student evaluation and grading in an agency different from the agency providing the instruction, as is done regularly in England and sporadically in the United States. External evaluating agencies almost invariably are concerned with summative rather than formative evaluation. This distinction is important partly as a way to emphasize the point that giving the task of summative evaluation to an external agency removes neither the opportunity nor the responsibility for evaluation from the teacher. Formative evaluation, which is the form most closely tied to the instructional process, remains a major responsibility of the teacher even when summative evaluation occurs externally.

When summative evaluation, with which grading is usually associated, is performed by an external agency, the competence of the students examined is certified according to some generally accepted standard. To the extent that grades are used outside the instructional institution, as in selection of graduates by other institutions or employers, the certification of an external agency might well replace grades (Goodman, 1964; Jencks and Riesman, 1968).

Placing the process of summative evaluation in an external agency does not necessarily remove grading entirely from the instructional institution. Just as other agencies may use the summative evaluations in selection, the instructional institution can use those evaluations for whatever internal purposes grades serve. These might include advancing students to higher-level courses, awarding

honors, encouraging promising students, or determining eligibility for extracurricular activities. For some purposes the external evaluations might be translated into grades within each institution, or within departments in an institution. This translation of external evaluations into internal grades might take account of the multitude of variables other than strict academic achievement that now enter the determination of grades—variables such as whether the course is part of the student's major field, the student's industriousness and attitudes toward the course, and the relative performance of other students in the course. These grades, with all their variation across situations, would then be distinct from those intended only to

indicate academic accomplishment and could be tailored to the specific purposes desired.

A major objection to external evaluation is its total dependence on limited observations of student performance conducted over a brief period of time. A student's instructors can almost certainly provide judgments about his capabilities that would not be duplicated by an examination, either internal or external. Although advantages and disadvantages can be found in each procedure, the choice between an external or internal examining agency depends heavily on the purpose of the examination and partly on the mundane issue of who should bear its cost.

VII. POSSIBLE DIRECTIONS FOR COLLEGE GRADING

Although college grading is currently the subject of widespread controversy, the points in hottest dispute are not fundamental issues. The liveliest issues today are how many grade categories to use, how to predict grades more accurately, and to a lesser extent, how to make grades comparable across courses, departments, and institutions.

The following issues that seem properly to merit prior consideration have been raised but not pursued. How well do grades serve the purposes for which they are intended? Do those purposes merit the enormous expenditures of time and energy grading entails? Would alternative ways of accomplishing the same purposes be preferable to current grading procedures? What are the unintended consequences of current grading practices, both for society at large and within the educational process? These issues involve external effects of the grading system.

Problems within the grading system influence external issues but can be pursued independently of them. The primary internal issue is that raised by Westland (1969). What do grades represent? The reasonably good internal consistency and short-term reliability that have been demonstrated indicate that grades in general, across varied courses and instructors, do reflect some common attribute. But that attribute can be called academic achievement, directed knowledge, verbal proficiency, academic facility, intellectual servility, or whatever is most commonly found to please professors. It is probably some complex entity in which several independent attributes are merged, as mass and volume are merged in density or as the height and weight of persons are merged in size. The most fruitful expenditure of effort with respect to the structure of grades would be that directed toward identifying the various components that underlie grades and assessing their interrelationships and fluctuations across fields, types of courses, professors, and students.

Knowledge of the various determinants of grades would then facilitate the study of external problems, such as

improving the selection and feedback processes. Grades could be made to reflect directly their various underlying dimensions, and selection procedures could be varied to suit the purposes of the selecting institution. Equivalence would no longer need to be forced onto inherently different measures. Prediction could probably be improved. Instructional goals could be more carefully defined and instructional effectiveness more adequately assessed.

A procedure such as that suggested by Elbow (1969) would provide the advantages of descriptive grading but could be carried out without many of the inconveniences presented by unsystematized prose statements in reporting achievement. A study of the current processes of student evaluation at an institution could reveal the most common dimensions of student performance that faculty members consider important in their own courses. An institution's evaluation of its students might be desirably broadened by including dimensions of performance found important in studies at other institutions, such as those by Hilton, Kendall, and Sprecher (1970) or Junius Davis (1964, 1965, 1966). Simple rating scales based on the desired dimensions could then constitute the basic achievement report. As Elbow suggests, faculty members could choose those dimensions they consider appropriate to describe a student's performance, remaining free, as at present, to determine the evaluation procedures on which their ratings would be based. Students might, as Feldmesser (1969) and Wolfle (1968) suggest, be involved in decisions as to which dimensions to include in the evaluations, and these could vary with different students. Averaging would occur only with respect to common dimensions and much of the richness of descriptive grading could be achieved without its administrative inconveniences.

An apparent deficiency in differentiated grading may be the tendency for a student's excellence in one area to color faculty judgments of his performance in other areas—the "halo effect." This tendency is equally present in any form

of grading but is more apparent when grading is differentiated into several dimensions. On the other hand, the explicitness of the various dimensions of performance may reduce the "halo effect" by making faculty members more aware of distinctions in types of performance.

One of the most important consequences of greater knowledge of grade components would be an increased likelihood of demonstrating connections between academic performance and behavior outside the academic setting. While academic growth, as represented by advancement through collegial institutions, may be defended as inherently desirable, it would gain public support, recognition, and understanding if its importance to nonacademic enterprises could be shown more convincingly. In a period of increasing calls for accountability in higher education, few issues seem more pertinent than a demonstration of what is meant by, implied in, or associated with the kind of academic growth that colleges claim good grades represent.

A major deficiency in current grading procedures is their broad uniformity in spite of the variety of functions demanded of them. An obvious direction for improvement would be to vary the form to suit the purpose. Some form of differentiated grading at midterm, for example, would probably serve the feedback function of grades far better than do present procedures. Whether these grades should be retained in a student's central record, only in the instructor's records, or not at all would depend on how well they were suited to purposes other than feedback to students and on the availability of other methods to accomplish those purposes.

For many of the administrative purposes within the institution, grades seem unnecessary other than as an indication that a student has completed a course satisfactorily. Eligibility for various activities, or for considerations such as veterans' benefits and other financial assistance, seems to justify no requirements other than bona fide status as a student. Whether one student is more or less capable than another has no obvious relevance to administrative considerations associated with status as a student. Pass/No Record grading would effectively serve this kind of administrative purpose.

Selection within the institution for academic awards, honor programs, or special classes could well be based on faculty nominations supported by evaluative information provided by the faculty. If this practice were followed, faculty members might retain their own records of differentiated grading reports to students. From these, informed nominations could readily be made. The nature of these purposes makes detailed information on all students unnecessary.

Evaluations of the effectiveness of different programs or departments or of new instructional procedures depend on summative evaluations of student performance. These could be carried out internally or by an external agency and could take a variety of forms, including comprehensive examinations, evaluation of student products, or evaluation of student portfolios accumulated during a course. A major

consideration in evaluations for these purposes is that all students need not be evaluated. The teaching effectiveness of any department can be adequately assessed by testing a sample of its students and testing individual students in the sample only partially.

If 300 students have taken a two-semester course in economics, for example, the effectiveness of that course sequence could be evaluated by having six groups of 20 students each take a 1-hour segment of a 6-hour comprehensive examination. The unexamined 180 students could be assigned to comprehensive examinations in other courses. Since the examinations would be used to assess courses and instruction rather than students, most of the anxiety associated with examinations should be avoided, at least in the students if not in the instructors. Assignment of students to examinations could be done randomly just before the exams are given so students would not study disproportionately in the area in which they were to be examined.

Certification of students' accomplishments in certain areas, when a more elaborate indication than satisfactory completion of a specified set of courses is desired, could be accomplished by some form of summative evaluation in which individual students were examined in all pertinent areas. Again, these could be internal or external examinations. Whether the instructional institution provided evaluative information for other agencies, such as graduate and professional schools, could be determined by each institution. A useful procedure might be for undergraduate and graduate institutions to plan jointly for a form of summative evaluation that would serve both institutions. Students uninterested in advanced education need not be subjected to that evaluative procedure.

A number of faculty members at the University of California at Berkeley favor the use of summative evaluation at the end of a multicourse sequence (*Education at Berkeley*, 1968). Grades would not be assigned in the early courses in the sequence, but performance in these courses would be reflected in the supergrade at the end of the sequence. Grading, in their view, would be improved by reducing its frequency and increasing its comprehensiveness. Raimi (1967) made a similar proposal but suggested that the periodic comprehensive examinations not be tied to any particular sequence of courses. These procedures are sound in terms of summative evaluation and its purposes. Formative evaluation would, as is almost always the case, be another matter.

The purposes grades serve need clearer identification and more intensive examination to justify the expenditure of resources for their accomplishment and to determine the most effective ways they can be accomplished. The current outmoded and largely ineffective grading procedures should be replaced with procedures more specifically directed to their intended purposes. More varied and more effective procedures are available. That they have been used so little may be due to uncertainty or confusion over what is really wanted of grades.

In summary, we don't know what present grades represent as indexes of academic performance. The current issues surrounding grades and grading cannot be effectively faced until we do. When the components and structure of

grades are better described, we will be able to attack not only the current, rather limited issues, but the more substantial ones that bear heavily on the entire higher educational enterprise.

REFERENCES

- Aiken, Lewis F., Jr. "The Grading Behavior of a College Faculty." *Educational and Psychological Measurement* 23, Summer 1963, pp. 319-322.
- . "Intra-Departmental Variability and Student Expectations of College Grades." *Educational and Psychological Measurement* 24, Winter 1964, pp. 823-829.
- Andershalter, O.F. "Developing Uniform Departmental Grading Standards in a University." *The Journal of Experimental Education* 31, 1962, pp. 210-211.
- Axelrod, Joseph. "What Do College Grades Mean? A Survey of Practice at Four Institutions." In H.A. Estrin and D.P. Goode, eds. *College and University Teaching*. Dubuque, Iowa: Brown, 1964, pp. 466-472.
- . "The Creative Student and the Grading System." In Paul Hiest, ed. *The Creative College: Student: An Unmet Challenge*. San Francisco: Jossey-Bass, 1968.
- Barritt, Loren Spencer. "Note: The Consistency of First Semester College Grade Point Average." *Journal of Educational Measurement* 3, Fall 1966, pp. 261-262.
- Bartlett, James W. "Medical School and Career Performances of Medical Students with Low Medical College Admission Test Scores." *Journal of Medical Education* 42, 1967, pp. 231-237.
- Becker, Howard S.; Geer, Blanche; and Hughes, Everett C. *Making the Grade: The Academic Side of College Life*. New York: Wiley, 1968.
- Becker, Howard S.; Geer, Blanche; Hughes, Everett C.; and Strauss, Anselm L. *Boys in White: Student Culture in a Medical School*. Chicago: University of Chicago Press, 1961.
- Benson, Winston W. "Graduate Grading Systems." Washington, D.C.: Paper presented at the Council of Graduate Schools in the United States, 1969, ED 036 262. MF - \$0.65, HC - \$3.29.
- Berdik, Ralph F. "A Solution to the Problem of Distributing Course Grades." *School and Society* 93, 1965, pp. 373-375.
- Birney, Robert C. "The Effects of Grades on Students." *The Journal of Higher Education* 35, February 1964, pp. 96-98.
- Bloom, Benjamin. "Learning for mastery." *Evaluation Comment* 1, May 1968.
- Boldt, Robert F. "Factor Analysis of Business School Grades." Research Bulletin RB-70-49. Princeton: Educational Testing Service, 1970.
- Bowers, John F. "A Test of Variation in Grading Standards." *Educational and Psychological Measurement* 27, Summer 1967, pp. 429-430.
- Bramer, George R. "Grading and Academic Justice." *Improving College and University Teaching* 18, Winter 1970, pp. 63-65.
- Brooks, Walter L. "A Study of Penalty Grading and Probation Practices at Shasta College." Redding, California: Shasta College, 1968, ED 074 376. MF - \$0.65, HC - \$3.23.
- Buchman, W.A. "Grading systems across the country . . ." Evanston, Ill.: Center for Curriculum Design, Kendall College, 1970.
- Bundy, Stuart M. "Prediction of Success in the Doctoral Program of the School of Education of the University of Southern California." *Disertation Abstracts* 28A, May 1968, pp. 4523-4524.
- Burwen, Leroy S. Verbal report of 1970 survey. San Francisco State College, San Francisco, Calif., 1970.
- Calhoun, R.P. and Reddy, A.C. "The Frantic Search for Predictors of Success." *Journal of College Placement* 28, February-March 1968, pp. 54 ff.
- Caplow, Theodore. *The Sociology of Work*. Minneapolis: University of Minnesota Press, 1954.
- Christin, P.E. "A Style for the Seventies." *Educational Record* 51, Spring 1970, pp. 174-181.
- Clark, Burton R. *Educating the Expert Society*. San Francisco: Chandler Publishing Co., 1962.
- Clark, Edward L. "Reliability of Grade Point Averages." *The Journal of Educational Research* 57, April 1964, pp. 428-430.
- Cele, William G. "Breaking the grade-and-credit mold." *The Challenge of Curricular Change*. New York: College Entrance Examination Board, 1966, pp. 42-50.
- Committee on Educational Policy. "Report on Grading at the University of California, Santa Cruz." Santa Cruz, California: University of California, 1970.
- Committee on Grading. "Final Report and Recommendations." Berkeley, California: School of Law, University of California, 1970.
- Croner, Ward. "An Empirical Investigation of Student Attitudes Toward the Pass-Fail Grading System at Wellesley College." Philadelphia: Paper presented at Eastern Psychological Association, 1969.
- Dale, Wesley. "Concerning Grading and Other Forms of Student Evaluation." Washington: Paper presented at the Council of Graduate Schools in the United States, 1969, ED 036 260. MF - \$0.65, HC - \$3.29.
- Davis, James A. "The Campus as a Frog Pond: An Application of the Theory of Relative Deprivation to Career Decisions of College Men." *The American Journal of Sociology* 72, July 1966, pp. 1-16.
- Davis, Junius A. "Faculty Perceptions of Students. II. Structure of Faculty Characterizations." Research Bulletin RB-64-12. Princeton: Educational Testing Service, 1964.
- . "What College Teachers Value in Students." *College Board Review*, Spring 1965, pp. 15-18.
- . "Faculty Perceptions of Students. VI. Characteristics of Students for Whom There is Faculty Agreement on Desirability." Research Bulletin RB-66-28. Princeton: Educational Testing Service, 1966.
- Dickenson, C. "What Employers Look For in the College Graduate." *The Personnel and Guidance Journal* 33, February 1955, pp. 460-464.
- Domino, George. "Differential Prediction of Academic Achievement in Conforming and Independent Settings." *Journal of Educational Psychology* 59, October 1968, pp. 256-263.
- Dressel, Paul L., and Nelson, C.H. "Testing and Grading Policies." In Paul L. Dressel and Associates. *Evaluation in Higher Education*. Boston: Houghton Mifflin, 1961, pp. 227-252.
- Dyer, Henry. "Practice at War with Purpose in Schools and Colleges." San Francisco: Talk given at the College Entrance

- Examination Board Western Regional Meeting. 1967.
- Ebel, Robert L. "Content Standard Test Scores." *Educational and Psychological Measurement* 22, Spring 1962, pp. 15-25.
- . *Measuring Educational Achievement*. Englewood Cliffs, N.J.: Prentice-Hall, 1965.
- . "The Relation of Scale Fineness to Grade Accuracy." *Journal of Educational Measurement* 6, Winter 1965, pp. 217-221.
- Education at Berkeley: Report of the Select Committee on Education*. Berkeley: University of California Press, 1968.
- Elbow, Peter H. "More Accurate Evaluation of Student Performance." *The Journal of Higher Education* 40, March 1969, pp. 219-230.
- Elton, L.R.B. "The Assessment of Students—A New Approach." *Universities Quarterly* 22, June 1968, pp. 291-301.
- Ericksen, Stanford C. "Grades." *Memo to the Faculty*, No. 18. Ann Arbor, Michigan: Center for Research on Learning and Teaching, The University of Michigan, 1966.
- . "Pass-Fail Grading." *Memo to the Faculty*, No. 22. Ann Arbor, Michigan: Center for Research on Learning and Teaching, The University of Michigan, 1967.
- Feldmesser, Robert A. *The Option: Analysis of an Educational Innovation*. Hanover, N.H.: Dartmouth College, 1969.
- Forlyce, J.W., and Bromley, A. "Santa Fe Junior College." *Junior College Journal* 40, December-January 1969-70, pp. 46-49.
- Freeman, James T. "A Summary Progress Report on an Experimental Study of a Pass/No Report Grading System." San Bernardino, California: California State College at San Bernardino, 1969.
- Fricke, Benno. "Evaluation of Student Achievement." *Memo to the Faculty*, No. 11. Ann Arbor, Michigan: Center for Research on Learning and Teaching, The University of Michigan, 1965.
- Friedenberg, Edgar Z. "The Real Functions of Educational Testing." *Change in Higher Education* 2, 1970, pp. 43-47.
- Funkenstein, D.H. "The Learning and Personal Development of Medical Students and the Recent Changes in Universities and Medical Schools." *Journal of Medical Education* 43, 1968, pp. 883-897.
- Garnson, Zella F. "Performance and Personalism in Student-Faculty Relations: Departments of Natural Sciences and Social Sciences." *Sociology of Education* 40, Fall 1967, pp. 279-301.
- Glazer, Nathan. "Are Academic Standards Obsolete?" *Change in Higher Education* 2, November-December 1970a, pp. 38-44.
- . "On Grading." In Nathan Glazer, *Remembering the Answers: Essays on the American Student Revolt*. New York: Basic Books, 1970b, pp. 245-249.
- Gohn, Lyle A. "An Investigation of the Selection Techniques of Veterinary Science and Medicine Students at Purdue University." *Dissertation Abstracts* 28A, May 1968, p. 4445. (Abstract)
- Gold, Martin. "A Letter from Prof. Martin Gold." *Memo to the Faculty*, No. 16. Ann Arbor, Michigan: Center for Research on Learning and Teaching, The University of Michigan, 1966.
- Goldstein, Kenneth M., and Tinker, Harvey A. "A Review of Grading System Practices in the Higher Education Institutions of New York State." Staten Island, New York: Richmond College, City University of New York, 1969.
- Goodman, Paul. "In What Ways Does the Present Marking and Credit System Inhibit or Promote Learning?" In *Current Issues in Higher Education: Undergraduate Education*, edited by G. Kerry Smith. Washington, D.C.: Association for Higher Education, 1964, pp. 123-125.
- Gough, Harrison G. "Nonintellective Factors in the Selection and Evaluation of Medical Students." *The Journal of Medical Education* 42, July 1967, pp. 642-650.
- Grant, J.L. "Justice in Grading." *Journal of Legal Education* 9, 1956, pp. 186-212.
- Haagen, C. Hess. "The Origins of a Grade." *The Journal of Higher Education* 35, February 1964, pp. 82-91.
- Hackman, J. Richard; Wiggins, Nancy; and Bass, Alan R. "Prediction of Long Term Success in Doctoral Work in Psychology." *Educational and Psychological Measurement* 30, Summer 1970, pp. 365-374.
- Haley, John V.; and Lerner, Melvin J. "The Dimensionality of Grades for the First Two Years of Medical School." *Journal of Medical Education*, 42, 1967, pp. 273-274.
- Hankon, L.W. "College Grades and Admission to Medical School." *The Journal of Higher Education* 35, February 1964, pp. 93-96.
- Hassler, William W. "Results of Pass-Fail Questionnaire Sent to Graduate School Deans." Indiana, Pennsylvania: Indiana University of Pennsylvania, 1969, ED 037 168. MF - \$0.65, HC - \$3.29.
- Heckman, Robert W.; Banas, Paul A.; Lazenby, Robert E.; and Moore, Loretta M. "Biographical Correlates of Management Success." *Proceedings of the 77th Annual Convention of the American Psychological Association* 4, 1969, pp. 611-612.
- Hepworth, Dean H. "The Predictive Validity of Admissions Criteria Used by the Graduate School of Social Work, University of Utah, to Select Students for the Fall Term, 1966." *Dissertation Abstracts* 29A, March 1969, pp. 3225-3226.
- Hills, John R., and Gladney, Marilyn B. "Factors Influencing College Grading Standards." *Journal of Educational Measurement* 5, Spring 1968, pp. 31-39.
- Hilton, Thomas L.; Kendall, Lorne, M.; and Sprecher, Thomas B. "Performance Criteria in Graduate Business Study: Parts I and II—Development of Rating Scales, Background Data Form and the Pilot Study." Research Bulletin RB-70-3. Princeton: Educational Testing Service, 1970.
- Holland, John L. "The Prediction of College Grades From Personality and Aptitude Variables." *Journal of Educational Psychology* 51, 1960, pp. 245-254.
- Horowitz, Milton J. *Educating Tomorrow's Doctors*. New York: Appleton-Century-Crofts, 1964.
- Hoyt, Donald P. "The Relationship Between College Grades and Adult Achievement: A Review of the Literature." *ACT Research Report*. No. 7. Iowa City: American College Testing Program, 1965.
- . "College Grades and Adult Accomplishment." *Educational Record* 47, Winter 1966, pp. 70-75.
- . "The Criterion Problem in Higher Education." In *Learning and the Professors*, edited by Ohmer Milton and Edward Joseph Shoben, Jr. Athens, Ohio: Ohio University Press, 1968.
- Humphreys, Lloyd G. "The Fleeting Nature of the Prediction of College Academic Success." *Journal of Educational Psychology* 59, October 1968, pp. 375-380.
- Husik, T.R. "Different Kinds of Evaluation and Their Implications for Test Development." *Evaluation Comment* 2, October 1969, pp. 8-10.
- Hyman, Lawrence W. "Grades and Academic Standards." New York: Brooklyn College, City University of New York, 1969, ED 036 275. MF - \$0.65, HC - \$3.29.
- Iadarola, Toni. "Grading the Graduate Student: A Student Perspective." Washington, D.C.: Paper read at the Council of Graduate Schools in the United States, 1969, ED 036 263. MF - \$0.65, HC - \$3.29.
- Jencks, Christopher, and Riesman, David. *The Academic Revolution*. New York: Doubleday, 1968.
- Johansson, Charles B.; Rossmann, Jack F.; and Sandell, Sandra D. "The S-U Grading Option." St. Paul, Minn.: Macalester College, 1970.
- Juola, Arvo E. "Illustrative Problems in College-Level Grading." *The Personnel and Guidance Journal* 47, September 1969, pp. 29-33.
- Kappel, F.R. "From the World of College to the World of Work." New York: American Telephone and Telegraph Co., 1952.
- Karlins, Marvin; Kaplan, Martin; and Stuart, William. "Academic Attitudes and Performance as a Function of Differential Grading Systems: An Evaluation of Princeton's Pass-Fail System." *The*

- Journal of Experimental Education* 37, 1969, pp. 38-50.
- Katz, Joseph, and Associates. *No Time for Youth: Growth and Constraint in College Students*. San Francisco: Jossey-Bass, 1968.
- Katz, Michael B. *The Irony of Early School Reform: Educational Innovation in Mid-nineteenth Century Massachusetts*. Cambridge, Mass.: Harvard University Press, 1968.
- Kelly, Samuel P., and Thompson, Ralph. "Grading and the Nature of the Discipline." *The Journal of Higher Education* 39, December 1968, pp. 517-518.
- Kinloch, Graham C. "Sponsored and Contest Mobility Among College Graduates: Measurement of the Relative Openness of a Social Structure." *Sociology of Education* 42, Fall 1969, pp. 350-367.
- Kirby, B.C. "Three Error Sources in College Grading." *The Journal of Experimental Education* 31, 1962, pp. 211-218.
- Klein, Stephen P., and Evans, Franklin R. "An Examination of the Validity of Nine Experimental Tests for Predicting Success in Law School." *Educational and Psychological Measurement* 28, Autumn 1968, pp. 909-913.
- Korn, Harold A. "Higher Education Programs and Student Development." *Review of Educational Research* 39, April 1969, pp. 115-171.
- Lacey, O.L. "How Fair Are Your Grades?" In *Higher Education in Engineering and Science*, edited by H.A. Estrin. New York: McGraw-Hill, 1963, pp. 348-352.
- Ladd, Dwight R. *Change in Educational Policy: Self-Studies in Selected Colleges and Universities*. New York: McGraw-Hill, 1970.
- Lannholm, Gerald V. "Review of Studies Employing GRE Scores in Predicting Success in Graduate Study 1952-1967." Graduate Record Examinations Special Report No. 68-1. Princeton: Educational Testing Service, 1968a.
- . "The Use of GRE Scores and Other Factors in Graduate School Admission." Graduate Record Examinations Special Report No. 68-4. Princeton: Educational Testing Service, 1968b.
- , Marco, Gary L., and Schrader, William B. "Cooperative Studies of Predicting Graduate School Success." Graduate Record Examinations Special Report No. 68-3. Princeton: Educational Testing Service, 1968.
- Lauter, Paul, and Howe, Florence. "How the School System is Rigged for Failure." *The New York Review of Books* 14, June 18, 1970, pp. 14-21.
- Lavin, David E. *The Prediction of Academic Performance: A Theoretical Analysis and Review of Research*. New York: Russell Sage Foundation, 1965.
- Law School Admission Test Council. "Statement on Pass-Fail Grading Systems." October 27, 1970, ED 045 043. MF - \$0.65, HC - \$3.29.
- Lewis, W.A., and Smith, W.C. "Factors Used in Grading and Their Relationships to Test Validation." *College Student Personnel Abstracts* 5, Spring 1970, pp. 331-332.
- Lunneborg, Clifford E., and Lunneborg, Patricia W. "The Prediction of Different Criteria of Law School Performance." *Educational and Psychological Measurement* 26, Winter 1966, pp. 935-944.
- Marshall, Max S. *Teaching Without Grades*. Corvallis, Oregon: Oregon State University Press, 1968.
- Mason, R.D. "Grades + Activities = Financial Success?" *Journal of College Placement* 25, April 1965, pp. 26 ff.
- Mayhew, Lewis B. "Contemporary College Students and the Curriculum." *SREB Research Monograph*, No. 14. Atlanta: Southern Regional Education Board, 1967.
- McClaine, R.E. "Factors in Executive Promotion and Demotion: An Empirical Study." *Dissertation Abstracts* 29A, 1968, p. 1326.
- Mehrabian, Albert. "Undergraduate Ability Factors in Relationship to Graduate Performance." *Educational and Psychological Measurement* 29, Summer 1969, pp. 409-419.
- Melville, George L., and Stamm, Eleanor. "The Pass-Fail System and the Change in the Accounting of Grades on Comprehensive Examinations at Knox College." Galesburg, Illinois: Office of Institutional Research, Knox College, 1967, ED 014 788. MF - \$0.65, HC - \$3.29.
- Midwest College Placement Association. "The Use of Personnel Tests and Biographical Data in the Selection of Recent College Graduates." West Lafayette, Indiana: The Association, 1964-65.
- Miller, Richard H. "Students Show a Preparation Increase But No Increase in Grades was Shown." *College and University* 45, Fall 1969, pp. 28-30.
- Miller, Stuart. *Measure, Number, and Weight: A Polemical Statement of the College Grading Problem*. Ann Arbor, Mich.: Center for Research on Learning and Teaching, the University of Michigan, 1967. Also Knoxville, Tennessee: Learning Research Center, The University of Tennessee, 1967.
- Milton, Ohmer. "Grades and Grading." *Teaching-Learning Issues*, No. 2. Knoxville, Tennessee: Learning Resources Center, The University of Tennessee, 1966.
- . "Pass-Fail Grading." *Teaching-Learning Issues*, No. 5. Knoxville, Tennessee: Learning Resources Center, The University of Tennessee, 1967.
- . "What It Is . . . I Measure I Do Not Know." *Educational Record* 49, Spring 1968, pp. 160-165.
- Morishima, James K., and Micck, Sidney S. "Pass-Fail Evaluation: Phase II: Questionnaire Analysis." Seattle: Office of Institutional Educational Research, University of Washington, 1970.
- Murphy, Lois B., and Raushenbush, Esther, eds. *Achievement in the College Years: A Record of Intellectual and Personal Growth*. New York: Harper, 1960.
- "Operation Opportunity (Program II) at Lake Forest College: A Report by the Preceptors." Lake Forest, Ill.: Lake Forest College, 1970.
- Paquette, J.W. "Student Credentials: What Elements Are Really Essential?" *Journal of College Placement* 26, April-May 1966, pp. 34 ff.
- Pass-Fail Study Committee. "Report to the Senate." Washington, D.C.: Phi Beta Kappa, 1969.
- Pemberton, W.A. "A Recommendation For Revision of Student Evaluation Procedures." Newark, Delaware: University of Delaware, 1969.
- Perry, Louis B. "College Grading: A Case Study and its Aftermath." *Educational Record* 49, Winter 1968, pp. 78-84.
- Puge, Fred L. "Teaching Effectiveness of 'A' and 'C' Elementary Teachers." *Journal of Educational Research* 62, November 1968, pp. 99-102.
- Pitcher, Anita, and Bosler, Robert. "Study of the 1969 UCSC Graduates." Santa Cruz, California: Office of the Chancellor, University of California, 1970.
- Pitcher, Barbara. "The Law School Admission Test as a Predictor of First-Year Law School Grades, 1962-1963." Statistical Report SR-65-32. Princeton: Educational Testing Service, 1965.
- Pitcher, Barbara; Dwyer, Alice B.; and Smith, Herman F. "An Exploratory Study of Four Variables as Moderators of the Effectiveness of the Admission Test for Graduate Study in Business and Undergraduate Record for Predicting Business School Grades." Statistical Report SR-68-65. Princeton: Educational Testing Service, 1968.
- Pitcher, Barbara, and Winterbottom, John A. "The Admission Test for Graduate Study in Business as a Predictor of First-Year Grades in Business School, 1962-1963." Statistical Report SR-65-21. Princeton: Educational Testing Service, 1965.
- Porter, Paul B. "Faculty Evaluation of Graduate Students." *Psychological Reports* 24, 1969, pp. 761-767.
- Priest, R.F. "Attitudes to Competitive Grading in College as a

- Function of Aspiration, Anxiety and Grades." Vancouver: Paper presented at the Western Psychological Association, 1969.
- Quann, C.J. "Pass-Fail Grading: What Are the Trends?" Paper presented at the American Association of Collegiate Registrars and Admissions Officers, 1970.
- Raimi, Ralph A. "Examinations and Grades in College." *AAUP Bulletin* 53, September 1967, pp. 309-317.
- "A Report on the Independent Study Program: 1965-1970." Claremont, Calif.: Pomona College, 1970.
- Richards, James M., Jr. *Assessing Student Performance in College*. Report 2. Washington: ERIC Clearinghouse on Higher Education, 1970.
- Riesman, David; Gusfield, Joseph; and Gamson, Zekia. *Academic Values and Mass Education: The Early Years of Oakland and Monteith*. New York: Doubleday, 1970.
- Roemer, R.E., S.J. "Nine Year Validity Study of Predictors of Medical School Success." *The Journal of Educational Research* 59, December 1965, pp. 183-185.
- Rosser, Marilyn. "Criteria Used for Admission by Graduate Departments." *American Psychologist* 25, June 1970, pp. 558-560.
- Rossmann, Jack E. "Graduate School Attitudes to S-U Grades." *Educational Record* 51, Summer 1970, pp. 310-313.
- Salyer, R.C. "Academic Performance in College and its Relationship to Successful Teaching." *Student Personnel Association for Teacher Education* 7, 1968-1969, pp. 43-48.
- San Francisco Sunday Examiner & Chronicle*, Section C, p. 2, January 10, 1971.
- Schrader, W.B., and Pitcher, Barbara. "The Advanced Tests of the Graduate Record Examinations as a Predictor of Law School Grades." *Law School Admission Test Annual Council Report*, 1970, pp. 108-129.
- Schwab, J.J. "Eros and Education: A Discussion of One Aspect of Discussion." *Journal of General Education* 8, October 1954, pp. 51-71.
- Scriven, Michael. "The Methodology of Evaluation." In *Perspectives of Curriculum Evaluation*, edited by Robert E. Stake. AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand-McNally, 1967, pp. 39-83.
- . "Discussion." In *Invitational Conference on Testing Problems*. Princeton: Educational Testing Service, 1969, pp. 112-115.
- Sexton, Patricia Cayo. *The American School: A Sociological Analysis*. Englewood Cliffs, N.J.: Prentice-Hall, 1967.
- Sgan, Mathew R. "The First Year of Pass-Fail at Brandeis University: A Report." *The Journal of Higher Education* 40, February 1969, pp. 135-144.
- Sharp, Laure M. *Education and Employment: The Early Careers of College Graduates*. Baltimore: Johns Hopkins University Press, 1970.
- Smallwood, Mary L. "An Historical Study of Examinations and Grading Systems in Early American Universities." *Harvard Studies in Education*, Vol. 24. Cambridge, Massachusetts: Harvard University Press, 1935.
- Smith, Lawrence G. "Non-Punitive Grading in California Junior Colleges." El Cajon, California: Grossmont College, 1969, ED 035 400. MF - \$0.65, HC - \$3.29.
- Sparks, David S. "Grading and Student Evaluation." Washington, D.C.: Paper presented at the Council of Graduate Schools in the United States, 1969. ED 036 261. MF - \$0.65, HC - \$3.29.
- Stake, Robert E. "Objectives, Priorities, and Other Judgment Data." *Review of Educational Research* 40, April 1970, pp. 181-212.
- Stallings, William M., and Leslie, Elwood K. "Student Attitudes Toward Grades and Grading." *Improving College and University Teaching* 18, Winter 1970, pp. 66-68.
- Stallings, William M.; Smock, H. Richard; and Leslie, Elwood K. "The Pass-Fail Grading Option." *School and Society* 96, March 16, 1968, pp. 179-180.
- Stewart-Tull, D.E.S. "The Setting and Marking of Microbiology Examinations." *Journal of Biological Education* 4, March 1970, pp. 25-42.
- Stordahl, Kalmer E. "Predicting Grades in a Master's Degree Program." *Journal of Educational Measurement* 4, Summer 1967, pp. 119-122.
- Teaf, H.M., Jr. "What Price Grades?" *The Journal of Higher Education* 35, February 1964, pp. 100-103.
- Thorndike, Robert L. "Marks and Marking Systems." In R.L. Ebel, ed. *Encyclopedia of Educational Research*. (4th ed.) Toronto: Macmillan, 1969, pp. 759-766.
- Torbert, William R., and Hackman, J. Richard. "Taking the Fun Out of Outfoxing the System: Yale University." In *The Changing College Classroom*, edited by Philip Runkel, Roger Harrison, and Margaret Runkel. San Francisco: Jossey-Bass, 1969, pp. 156-181.
- Trow, William Clark. "Grades and Objectives in Higher Education." *Educational Record* 49, Winter 1968, pp. 85-91.
- Tyler, Ralph W. "The Changing Structure of American Institutions of Higher Education." In *The Economics and Financing of Higher Education in the United States: A Compendium of Papers Submitted to the Joint Economic Committee, Congress of the United States*. Washington: Government Printing Office, 1969, pp. 305-320.
- Wallace, Walter L. *Student Culture: Social Structure and Continuity in Liberal Arts College*. Chicago: Aldine, 1966.
- Walters, R.W., Jr., and Bray, Douglas W. "Today's Search for Tomorrow's Leaders." *Journal of College Placement* 23, 1963, pp. 22 ff.
- Webb, Sam C. "Measured Changes in College Grading Standards." *College Board Review*, No. 39, 1959, pp. 27-30.
- Werts, Charles E. "Path Analysis: Testimonial of a Proselyte." *The American Journal of Sociology* 73, January 1968, pp. 509-512.
- , and Watkey, Donovan J. "Determinants of Changes in Career Plans During College." *Sociology of Education* 41, Fall 1968, pp. 401-405.
- , and ——. "A Student's Dilemma: Big Fish - Little Pond or Little Fish - Big Pond." *Journal of Counseling Psychology* 16, January 1969, pp. 14-19.
- Westland, Gordon. "The Philosophy of Student Assessment." *Universities Quarterly* 23, Summer 1969, pp. 350-360.
- Wharton, Keith. "P-N and Non-P-N Students: Reasons for Choosing the Option: Graduate School Plans." Minneapolis: Bureau of Institutional Research, University of Minnesota, 1969.
- Wiggins, Nancy; Blackburn, Margaret; and Hackman, J. Richard. "Prediction of First-Year Graduate Success in Psychology: Peer Ratings." *The Journal of Educational Research* 63, October 1969, pp. 81-85.
- Woffle, Dacl. "Are grades necessary?" *Science* 161, September 20, 1968, p. 1203.
- Woodring, Paul. *The Higher Learning in America: A Reassessment*. New York: McGraw-Hill, 1968.