INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in “sectioning” the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from “photographs” if essential to the understanding of the dissertation. Silver prints of “photographs” may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

University Microfilms International
300 North Zeeb Road
Ann Arbor, Michigan 48106 USA
St. John’s Road, Tyler’s Green
High Wycombe, Bucks, England HP10 8HR
WASSERMAN, James Francis, 1942-
GRADING IN HIGHER EDUCATION: A REVIEW
OF THE LITERATURE AND STUDY OF STUDENT
JUDGMENT CONCERNING OBJECTIVES OF GRADING
AND EFFECTIVENESS OF GRADING SYSTEMS.

The Ohio State University, Ph.D., 1976
Education, higher

Xerox University Microfilms, Ann Arbor, Michigan 48106
GRADING IN HIGHER EDUCATION: A REVIEW OF THE LITERATURE
AND STUDY OF STUDENT JUDGMENT CONCERNING OBJECTIVES OF
GRADING AND EFFECTIVENESS OF GRADING SYSTEMS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

James F. Wasserman, B.A., M.A.

*****

The Ohio State University

1976

Reading Committee:

Donald P. Anderson
Robert Lange
Robert Sutton

Approved By

[Signature]
Adviser
Faculty of
Educational Administration
Without an understanding and giving partner this dissertation would not have been completed. I gratefully acknowledge my wife, Barbara, who in addition to providing encouragement and accepting the required sacrifices typed several versions of the dissertation. I also wish to acknowledge my two young daughters, Julie and Susan, whose simultaneous arrival several months ago provided the final stimulus for the completion of the dissertation.

Thanks are due to Dr. Donald Anderson who directed the dissertation for his encouragement, understanding, and assistance to a student working away from the campus with the resulting complications and delays. Thanks also are due to Dr. Robert Lange whose help was invaluable and to Dr. Robert Sutton for his willingness to serve on short notice. Appreciation is also noted for the faculty of the College of Education for all they have given to me.

Finally, thanks are due to Sister Marie Emmanuel, S.C. who edited the final copy and to Mrs. Beverly Jordan who typed the final copy.
VITA

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1942</td>
<td>Born - Toledo, Ohio</td>
</tr>
<tr>
<td>1964</td>
<td>B.A., St. Meinrad College, St. Meinrad, Indiana</td>
</tr>
<tr>
<td>1965</td>
<td>M.A., Athenaeum of Ohio, Cincinnati, Ohio</td>
</tr>
<tr>
<td>1965-1967</td>
<td>Instructor, Department of Philosophy, Edgecliff College, Cincinnati, Ohio</td>
</tr>
<tr>
<td>1969</td>
<td>M.A., Xavier University, Cincinnati, Ohio</td>
</tr>
<tr>
<td>1969-1973</td>
<td>Assistant Academic Dean, College of Mount St. Joseph, Cincinnati, Ohio</td>
</tr>
<tr>
<td>1973</td>
<td>Academic Dean, College of Mount St. Joseph, Cincinnati, Ohio</td>
</tr>
</tbody>
</table>

FIELDS OF STUDY

Major Field: Higher Education Administration

Minor Fields: Educational Planning and Development
             Business Administration
TABLE OF CONTENTS

ACKNOWLEDGMENTS................................................................. ii
VITA................................................................. iii
LIST OF TABLES................................................................. vi
LIST OF FIGURES................................................................. viii
Chapter
I. INTRODUCTION, THE PROBLEM, HYPOTHESES.............. 1
  Statement of the Problem.............................. 5
  Limitations of the Study.......................... 6
  Hypotheses................................. 7
  Organization of the Remainder of the
  Dissertation........................................ 8
II. REVIEW OF THE LITERATURE................................. 9
  Brief History of Grading.......................... 10
  Grading Systems: Actual and Proposed...... 26
  Purposes or Objectives of Grading........... 71
  Determinants of Grades......................... 76
  Grading Outcomes.................................. 90
  Technical/Theoretical Issues................. 131
  Summary........................................... 190
III. EXPERIMENTAL PROCEDURE............................. 192
  Introduction and Background.................. 192
  Research Design.................................. 195
  Data Tabulation.................................. 200
  Data Analysis..................................... 201
IV. Findings................................................................. 202
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypotheses Concerning the Objectives of Grading</td>
<td>202</td>
</tr>
<tr>
<td>Hypotheses Concerning Grading Systems</td>
<td>205</td>
</tr>
<tr>
<td>Analysis of Ratings by Grade Point Average</td>
<td>220</td>
</tr>
<tr>
<td>Summary of Findings</td>
<td>222</td>
</tr>
<tr>
<td>V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS</td>
<td>228</td>
</tr>
<tr>
<td>Background for the Study</td>
<td>228</td>
</tr>
<tr>
<td>The Study Methodology</td>
<td>228</td>
</tr>
<tr>
<td>Summary of Findings and Conclusions</td>
<td>229</td>
</tr>
<tr>
<td>Summary of Grading Literature</td>
<td>233</td>
</tr>
<tr>
<td>Discussion and Recommendations</td>
<td>244</td>
</tr>
<tr>
<td>LIST OF REFERENCES</td>
<td>255</td>
</tr>
<tr>
<td>APPENDIX</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>293</td>
</tr>
<tr>
<td>B</td>
<td>300</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table                                                                 Page
1. Summary of Analysis of Variance of Rating of Importance of Grading Objectives by Underclassmen, Upperclassmen--Arts and Sciences, Upperclassmen--Professional 204
2. Newman-Keuls Analysis of Overall Main Effect of Rating of Grading Objectives 206
3. Summary of Analysis of Variance of Utility Values of Grading Systems by Underclassmen, Upperclassmen--Arts and Sciences, Upperclassmen--Professional 209
4. Summary of Analysis of Variance for Simple Main Effects of Grading Systems on Utility Scores of Grading Systems 211
5. Newman-Keuls Analysis of Simple Main Effect of Rating of Grading Systems by Underclassmen 212
6. Newman-Keuls Analysis of Simple Main Effect of Rating of Grading Systems by Arts and Science Students 213
7. Newman-Keuls Analysis of Simple Main Effect of Rating of Grading Systems by Professional Students 214
8. Summary of Analysis of Variance for Simple Main Effects of Groups on Utility Values of Grading Systems 217
Table

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Chi Square Analyses of Student Ratings of Grading Objectives According to Grade Point Average Intervals</td>
<td>221</td>
</tr>
<tr>
<td>12. Chi Square Analyses of Student Ratings of Grading Systems According to Grade Point Intervals</td>
<td>223</td>
</tr>
<tr>
<td>13. Chi Square Table for Descriptive Grading System</td>
<td>224</td>
</tr>
<tr>
<td>14. Chi Square Table for A-F With + and - Grading System</td>
<td>225</td>
</tr>
<tr>
<td>15. Student Rating Scores for Importance of Grading Objectives--Pilot Study</td>
<td>299</td>
</tr>
<tr>
<td>17. Design Layout for Student Rating Scores for Importance of Grading Objectives</td>
<td>300</td>
</tr>
<tr>
<td>18. Design Layout for Student Rating Composite Summary Scores for Effectiveness of Grading Systems</td>
<td>301</td>
</tr>
<tr>
<td>19. Group Means for Rating of Objectives</td>
<td>302</td>
</tr>
<tr>
<td>20. Group Means for Utility Scores of Grading Systems</td>
<td>303</td>
</tr>
<tr>
<td>21. Mean Ratings of Grading Objectives by Students According to Grade Point Average Intervals</td>
<td>304</td>
</tr>
<tr>
<td>22. Mean Ratings of Grading Systems by Students According to Grade Point Average Intervals</td>
<td>305</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>1. Mean Group Rating of Grading Objectives</td>
<td>203</td>
</tr>
<tr>
<td>2. Mean Ratings of Grading Objectives by Three Student Groups</td>
<td>207</td>
</tr>
<tr>
<td>3. Mean Ratings of Grading Systems by Three Student Groups</td>
<td>210</td>
</tr>
<tr>
<td>4. Mean Ratings of Grading Systems by Three Student Groups by Six Grading Systems</td>
<td>216</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION, THE PROBLEM, HYPOTHESES

The grading system used in undergraduate higher education has evolved approximately over the last one hundred years. It is analogous to an accounting system, for student academic achievement is based upon credits, courses, grades, and the grade-point-average (GPA). Because of the quantification according to both quality and magnitude, decision making can take place according to pre-determined rules, specified quantities of credits at specified grade levels, for example, being the basis for decisions concerning continuance of study and graduation. Convertibility and interchangeability in grading data are provided between colleges and graduate schools and employers are permitted to use undergraduate records for various selection procedures. To summarize the complex grading situation, one may say that as higher education became mass education growing numbers of persons have been involved in the grading system, until that system is now truly an integral part of the larger society.

Because grades and grading occupy this major role today, they have become central to the teaching-learning process in important ways. A student's success both in and after school, for example, is defined in terms of grades, so they become salient motives in his learning behavior. Consequently, a student may choose or reject a course, a major, or a school according to his perception of potential grade or grades. For the same reason a student, depending upon the evaluation used by an instructor to determine a grade, may undertake serious and significant study, or concentrate on memorizing trivia. The teaching method and the philosophy of grading of the instructor are closely related and students as a result tend to evaluate faculty in accord with the established grading criteria and their consistency in application.
Serious problems surround the entire grading area, and many prominent and respected educators have delineated their concerns. A graduate dean, for example, pointed out the ambiguity in meaning of grades, the distortion by them of the learning process, and the questionable uses made of them.1 Another educator stressed the extrinsic motivation provided by grades, the negative student-teacher relations, and the failures resulting from the marking system.2 Still another was disturbed by the tendency of grades to become ends in themselves, by their inflexibility, by the motivation engendered by them for students to beat the statistics of the system and the idiosyncracies of professors, rather than to learn.3 Of general concern also is the fact that since many important educational goals, such as freedom from irrational prejudice, independence of judgment, critical thinking and creativity are very difficult to measure on the usual sort of test, that students perhaps are being "taught" to ignore them.

Grades do fulfill a variety of functions or objectives, as suggested above, functions cited by defenders of traditional grading, including: (1) selection for employers and graduate schools; (2) motivation for the student to study; (3) guidance—differential marks guide the student to and from areas of greater and lesser competence; and (4) instruction—grades provide feedback on progress to the student and to the instructor.4

The objectives of grading, however, are also a matter of serious dispute. One former use of grades—by the Selective Service—is generally considered unacceptable today.


Nevertheless, other functions for grades, such as selection and motivation, are strongly defended. Understanding and agreement on the objectives of grading are fundamental requirements to resolve all other grading issues. Yet fundamental disagreement about the very objectives of grading exists, despite the critical importance of grading in undergraduate education, and though the thoughtful criticism of significant educators has been extensive, the literature indicates that little substantive research has been directed to the area.5

In the most complete review of grading to date, conducted by Warren and including approximately 200 articles published from 1965 to 1970, the author commented:

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 ... have been examined while other, more inclusive questions have been totally ignored.

About 100 of the studies dealt with pass-fail grading and with the prediction of grades in graduate or professional schools. Warren concluded:

These reports, in spite of their variety, leave large gaps in our knowledge about grades and grading. They lead only to a few general statements that can be made with much confidence ...


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.7

Since no comprehensive review of the literature of grading has been completed, with the exception of the 1965-1970 literature compiled by Warren and some restricted reviews of special topic areas, much of the information to be derived from it is practically unavailable. Coincident with, if not directly a result of, the unavailability of the literature, is the tendency of writers to perpetually argue the same issues, to produce the same descriptive studies, and to rewrite the same genre of opinion articles.

A serious gap has also existed in the failure to take into serious account input from students, since available grading literature primarily reflects the viewpoint of the system as seen by the institution and the instructor. True, superficial polls or questionnaires have been directed at students and have been reported, but these studies at one time indicated an almost complete docility of students to the grading system, and at another time showed that a very large minority wished to abolish all grades. Very little valid information has been obtained relating to student thought on deeper aspects of the problem, especially on topics like grading objectives or grading methods, which might facilitate intelligent discussion about grading. Faculty as a result do not understand student thought and values, or see how the system, of which the faculty are a part, orients student thought and values in undesired ways. And instructors even as they respond to questions from grade-oriented students about the specific content of the next examination continue to wish that students were more concerned about learning for the sake of learning.

7Ibid., p. 2.
Statement of the Problem

Significant problems exist in present grading practices. Educators seriously question whether the dysfunctions of grades outweigh their supposed objectives, and many disagree about what those objectives should be. As a result it is difficult even to find the basis for initiating rational discussion about grading, much less to implement better grading systems. Furthermore, the literature of grading is not helpful, because much of it is not available in useful form and because it tends to focus on narrow areas. Reviewing it leads one to conclude that little significant empirical research and less experimental research on the subject has been conducted. One serious gap in the literature occurs in the absence of significant input from students.

Accordingly it is the purpose of this study to provide a comprehensive review of the literature of grading which will be useful in making available what significant information has been collected, in serving as a guide for future research on grading, and in providing, if possible, some theoretical context for the consideration of grades. Specifically the review considers the history of grading to provide a valuable perspective for viewing present problems. It explores actual and theoretical grading systems appearing in the literature in order to clarify grading options, to provide a basis for comparison and theoretical discussions, and to select those grading systems to be used in the empirical study. The overriding importance of the objectives of grading, already noted, occupies a major part of grading literature and is critical to all subsequent issues. The review develops the major points of view concerning grading objectives and selects those of greatest significance to be used in the empirical study. The review also considers the effects of grading upon students in respect to learning and other outcomes in order to compare them both to grading objectives and results reported by students in the study. Finally, technical and theoretical concerns are reviewed to provide a context for interpreting the results of the study and as a basis for recommendations concerning grading.

Secondly, the study will attempt empirically to focus upon student input in the areas of the objectives of grading and of the effectiveness of various actual and theoretically significant suggested grading systems in achieving these objectives. Students will judge the suitability of
various grading objectives and the effectiveness of various grading systems significant in the literature. Several categories of students, such as underclassmen and upperclassmen will be used to determine the importance of such categories on student opinion.

The specific objectives of the study are:

1. From the literature to determine:
   a. Actual grading practices.
   b. Proposed functions or decision-making objectives served by grades.
   c. Effectiveness of grades in serving these objectives.

2. From student opinion to determine:
   a. Relative importance or desirability of the proposed functions or decision-making objectives supposed to be served by grades.
   b. Overall measure of the utility of various grading systems in attaining these functions.
   c. Differences in judgment among selected groups of students.

Limitations of the Study

The study is designed to provide a framework and perspective for studying and planning grading systems, in particular from the viewpoint of the student. It is not expected to lead to any "definitive" or "best" grading system.

The study is not an "experimental" study, in the sense of being an experimental design and statistical study which compares the real effects of various grading systems, though it may, however, as a result, contribute to the development of that kind of study.

It is probably true that no one system of grading is best for all situations, given the complexity of grading problems themselves, the unique conditions of particular institutions, the variety of teaching methods, the difference in classes and in areas of study, and especially by the uniqueness of the individual student who acts for a variety of motives.
Also, the care with which the individual instructor plans the learning process and the evaluation of student achievement is more important than the manner for reporting grades designed by the institution. Given adequate care in this area and close personal contact with the students being evaluated, any grading system might be adequate. Without this, none can be so.

Finally, the conclusions reached in the study will be general conclusions. The needs of individual institutions with enrollments of such disparity as one thousand to forty thousand and the practical and logistical considerations of grading will not be part of the study.

**Hypotheses**

The study was designed to test the following two sets of hypotheses:

Concerning the objectives of grading:

1. That grading objectives are most accepted by majors in professional areas and least accepted by underclassmen.

2. That certain objectives, such as the selection functions and mark of recognition for achievement, will receive more acceptance than other objectives, such as means of rewarding desired behavior and providing motivation.

3. That certain objectives will be more acceptable to one group of students than to the others, specifically:
   a. Underclassmen will consider motivation and mark of recognition, particularly undesirable as grading objectives.
   b. Students in professional areas will consider selection for graduate school and employer and means of rewarding desired behavior, particularly desirable as grading objectives.
Concerning grading systems:

1. That the more traditional grading systems will receive the greater acceptance.

2. That grading systems will be more accepted by students in professional fields and less accepted by underclassmen.

3. That more traditional grading systems will be most accepted by students in professional areas, and that less traditional grading systems will be most accepted by underclassmen.

Organization of the Remainder of the Dissertation

The additional chapters are titled Review of the Literature, Experimental Procedure, Findings, and Discussion and Recommendations, respectively. The review of the literature was comprehensive and attempted to include all significant work related to undergraduate grading practices, as well as occasional writing of a specific genre and literature from other fields which provide appropriate background. The review includes a brief historical review of grading, a consideration of various grading systems with a detailed review of pass-fail, and a study of the functions of grades, of the determinants of grades, of grading outcomes and of other technical and theoretical matters relating to grades.

The chapter on procedures includes discussions of the sample, the questionnaire instrument, and a description of a pilot study.
CHAPTER II

REVIEW OF THE LITERATURE

The literature of grading in American institutions of higher education spans more than two centuries. Much of the very early literature is found in college catalogs and chronicles of the history of individual institutions, some of which have been researched by Mary Smallwood in a study of grading practices in the nation's first institutions of higher education.\(^1\) Summaries and reviews of later grading literature have appeared occasionally in Review of Educational Research and other sources, among them a significant recent survey by Jonathan Warren, which reviewed approximately 200 articles appearing from 1965 to 1970.\(^2\)

The survey of the literature presented in this chapter has attempted to evaluate the significant literature which would be relevant in establishing a perspective and basis for understanding, interpreting, and making judgments about the current issues raised in regard to grading and the results of the empirical study. It also attempted to specify significant grading objectives and important grading systems in preparation for the study, and to ascertain grading outcomes as a basis for comparison with and interpretation of the results of the study.

The organization of the review generally proceeds from empirical to theoretical concerns. A brief history of grading is followed by a study of actual and proposed grading systems, of grading in practice, of determinants of

---


grades, of student and faculty grade orientations or perspectives, of side-effects of grading, and of the relationship of grading to other variables. The consideration of theoretical concerns, such as the precision and reliability of grades, the functions or objectives of grading, and grading in the context of evaluation and learning systems follows.

It is difficult to maintain strict boundaries between topics, since the areas by their nature are interrelated; a particular grading system, for example, will be influenced by assumed grading functions, and so on. Consequently there is frequent cross-reference between topics.

**Brief History of Grading**

Concern for the evaluation and recording of student achievement, as noted, has long been familiar in the world of higher education. Colonial colleges wrestled with many of the problems and concerns still perplexing colleges today, as can be seen by the glimpse into this era provided by Smallwood in *An Historical Study of Examinations and Grading Systems in Early American Universities*.

The primary measure of student achievement in the Colonial period took the form of an oral examination administered at wide intervals by a panel composed of the president, tutors, and other interested parties. By 1840, this method began to fall into disuse because of the inherent unreliability of oral examinations and the tendency of some tutors to feed the students easy questions so that a brilliant answer might reflect upon a brilliant instructor. The method also did not allow for reliable comparisons of achievement between students, since every student had to respond to different questions.

By the middle of the Nineteenth Century, what was at the time considered a major advance, namely, written examinations, had come into wide usage, permitting somewhat greater objectivity, in that all students responded to the same questions, and that these responses could be evaluated by any examiner. Some critics suggested that the daily recorded

---

3 Smallwood, *Examinations and Grading Systems*. 
grades for recitation provided a better measure of achievement than widely-spaced examinations, but advocates of written examinations countered that comprehensive written examinations were valuable in leading students to integrate the subject matter of the curriculum. In a further study, President Eliot of Harvard went a step beyond that, asserting that the teaching and the examining functions ought to be performed by different persons, so that an adequate measure of learning as well as of teaching might be obtained.4

Accordingly, in response to general dissatisfaction, numerous attempts were made to revise and change methods of administering examinations. In 1883 the University of Michigan even tried to do away with them, discontinuing the general examinations at the end of the academic term, but making provision for each instructor to give examinations in his classes.5

The justification given for insistence on examinations generally was of two sorts. Smallwood noted:

The purposes of examinations fall under two headings: the first for the purpose of measuring individual attainment; the second from the point of view of motivation. The latter is often filled by an occasional attempt at justification or rationalization. Along this line there is much similarity in the institutions studied.6

According to Smallwood, therefore, the function of examinations was twofold: to measure attainment and to spur the student on to more and better work. That recognition of student achievement, she noted, has taken several forms. In Colonial times a prized oratorical assignment was made to an outstanding student, who might be asked to deliver an oration at commencement. After the Colonial Period, more modern grading systems were introduced; for instance, at Yale President Stiles separated his students into four categories:

4Ibid., p. 22.
5Ibid., p. 22.
6Ibid., p. 22.
"optimi, second optimi, inferiores (but nonetheless boni), and peiores—which translate as best, next best, satisfactory, and inferior."\(^7\) In the same period, students were also ranked in four categories at William and Mary: "(1) first in their class, orderly and attentive, and making flattering improvement; (2) orderly and attentive, and making respectable improvement; (3) making little improvement, probably from want of diligence; and (4) having learned little or nothing, probably on account of escapades and idleness."\(^8\)

Over a century later, in 1911, a William Foster included this observation in his review of grading practices:

A century ago a Virginia academy attempted to have its students graded in six divisions, --bonus, melior, optimus, and malus, peijor, pessimus. But history records that 'the continual tendency was to mark inferior students too high. Thus it came to pass that not half the bad scholars got malus, the worst almost never fell below it, and bonus, though a mark of approbation, came to be considered as a disgrace, while optimus, which ought to have been reserved for scholars of the highest merit, was commonly bestowed on all who rose above mediocrity.'\(^9\)

Numerical ratings followed the descriptive ratings. Smallwood reported:

The leaders in educational institutions after 1800 came to an acute realization that evaluation is a necessary and desirable part of academic procedure. As a result a variety of scales came into existence.\(^10\)

The number of intervals on scales originally adopted ranged widely from four units, to eight units, to twenty units and

\(^7\)Ibid., p. 42.

\(^8\)Ibid., p. 44.


\(^10\)Smallwood, Examinations and Grading Systems, p. 42.
upwards, at different colleges, but in the second half of the Nineteenth Century marking based upon 100 was most popular. For example, Michigan, which in 1851 used a pass-fail system, by 1864 had changed to a system based on a hundred units. However, by the turn of the century, some colleges were questioning the precision of the implied accuracy of these numbers and began to use the first five letters of the alphabet to distinguish levels of achievement.

In 1884, Harvard adopted a five-rank system as a replacement for the percentage system, with an expression of confidence by the faculty that this would be sufficient to judge academic achievement and reduce the stress on college rank and competition for marks. It was argued that different instructors had used different standards for grading, and since a student's rank was determined by the total of his marks, competition for college honors under the percentage system might lead students to choose courses because they seemed to promise easy marks rather than because of the educational value of the instruction involved.

In 1895, the five-rank system was abandoned in turn for a scale of merit consisting of "Failed," "Passed," and "Passed with Distinction," and a few years later Harvard again modified the system by adding pluses and minuses to qualify these expressions. By 1900 the A, B, C, D, F system was in general use.

In the Eighteenth and Nineteenth Centuries, grades were frequently accepted as a composite assessment of moral as well as of academic achievement, as is clearly evident in some of the grade descriptions cited above. Grades of students who did not adhere closely to norms in some colleges were lowered in proportion to the number of demerits received. Scales of academic achievement and merit and demerit were frequently kept together, so that a punishment for misbehavior resulted in a reduction of previously earned

---

11Ibid., pp. 48-49.
12Ibid., p. 84.
13Ibid., pp. 51-52.
Demerits which could affect scholastic standing were given for such offenses as absence from prayer, lectures, or other exercises; for unprepared lessons, for misconduct. In 1874 Harvard decided to introduce a new system of keeping separate records for scholarship and conduct.

The relationship of grades and ranking to motivation, which Smallwood had noted, was also a subject of argument in the early days of higher education. At Dartmouth, scholastic distinctions were abolished for a time by President Lord to insure that growth in virtue and not academic achievement would be the main objective in student motivation. Eliot, on the other hand, believed that students needed to be motivated in as many ways as possible, and held that the ranking of achievement by the college, since it offered additional incentive, should be retained.

The advent of the elective system around 1870 had a significant and inevitable impact upon grading because for the first time students could choose from a variety of courses, many newly introduced as additions to traditional programs, and were not obliged to follow a single-track curriculum. This system made it imperative that schools establish single-course examinations with limited objectives, in place of the comprehensive examinations covering subject matter studied by all students, so with few exceptions, the widely-spaced comprehensive examinations based upon a common curriculum were phased out. In his study of comprehensive examinations Jones noted:

Here in America the development of examinations has been an interesting one, for colleges started with a general examination at the end of the course, based largely on the classics, and were forced to drop it as

14 Ibid., p. 70.
15 Ibid., p. 74.
16 Ibid., p. 74.
differentiated curricula and the inroads of the elective system deprived administrators of any common basis of testing. Instead there grew up a strange regime which exalted near objectives—the passing of factual tests on small areas of knowledge—until graduation came to depend on a series of small efforts distributed over a long period.18

Colleges used grades for a variety of purposes during the Nineteenth Century including the determination of "class positions and honors, commencement roles, special awards and class divisions, awareness of individual differences, aggregates and weighting marks."19 Degrees with various distinctions (cum laude, magna cum laude, summa cum laude, for example) were awarded to publicly recognize academic achievement and scholarships were frequently given on the basis of academic rank as determined by grades. A minimum numerical average was also specified as a requirement for graduation. As Smallwood noted:

The faculties constantly had before them the idea that they must set a value on a person's demonstration of ability. The difficulty came in agreeing on any one way of expressing or measuring the values that were set.20

A primary purpose of grades, of course, has consistently been to maintain standards, and much of the policy concerning grades has resulted from the attempt to establish these standards. Implicit in examinations and grading is the need for some set of standards of attainment; out of this need has grown a number of efforts to determine and uphold the meaning of grades. In 1853, for instance, when one instructor at Harvard gave out too many high marks, the Regent was given authority to reduce them.21 There was a general belief that strictness and thoroughness in grading were required for real scholarship.


19 Smallwood, Examinations and Grading Systems, p. 54.

20 Ibid., p. 70.

21 Ibid., p. 78.
By way of maintaining standards, students with low grades were not presented as candidates for degrees. After numerical values came into use, colleges established various regulations which specified the minimum final average required for receiving a degree. They also gradually developed standards of educational values which were conceived in terms of examinations and grades. Smallwood reported:

All the colleges held what they conceived to be high aims, and examinations were held to be the measure of the student's success in attaining the standards.  

Various techniques were used which might be considered some form of grading. Rudolph, in his history of American higher education, noted that after the Civil War the climate on most campuses was not favorable to scholarship. Consequently, in the 1860's to 1880's, artificial stimulation was tried to increase motivation and large scale prize programs were initiated; in 1871, Williams College awarded a prize to the student with the most prizes.

According to Brubacher and Rudy, examinations and grades in the Twentieth Century continued to be powerful motivators, though extrinsic, despite all attempts to provide primary motivation. Practices which developed during the early Twentieth Century were the use of the objective test, grading according to a normal distribution, and separation of the teaching and examining functions at some colleges to encourage the student to study the subject matter rather than the particular teaching characteristics of the instructor.

During the Twentieth Century, two broad phases in the historical development were distinguished by Smith and Dobbin in their 1960 review of marks and marking systems. From 1910 to 1940, research interest focused primarily on the mechanical and semantic problems of grading, and from 1940 to 1960, they noted, there was more interest in improvement of grades in comprehensiveness and communication. Between 1910 and 1920,

---

22Ibid., p. 86.

23Rudolph, American College and University, pp. 287-306.

considerable concern was again aroused by studies disclosing the variability and unreliability of grades. At this time, the percentage (or letter grade equivalent of a percentage range) was in general use, with a trend toward use of a 3 to 7 point system. In the effort to secure reliability, objective tests became widespread and greatly affected the testing function: the scientific measurement movement was underway.25

A book by Foster in 1911 is one of the best examples of the concern of the time. Reviewing the grading policies and practices of a number of institutions while focusing on Harvard, Foster noted that in some nineteen elementary courses at Harvard, the percent of A's ranged from 35% in Greek to 1% in English. The percent of E's or failures ranged from 21% in engineering to 0.5% in botany.26 He also found that less complete data suggested even greater variation among intermediate and advanced level courses, and that the grades given in advanced courses were considerably higher than those in elementary courses, while intermediate courses occupied a middle position. He also reported that similar patterns existed at other colleges, and that individual variation by instructors was greater than that by courses.27

Foster rejected the rationale that higher grades occurred in some courses because students had greater ability, and indicated that the opposite was true, that poorer students tended to choose courses where higher grades were given. As a remedy, a "scientific" distribution of grades on a modified normal curve was recommended, where grades should follow a distribution determined by the faculty. It was understood that some variations might occur in the short, but not in the long run, and wide deviations would have to be justified by the individual professor.28 Foster cited the
26Foster, College Curriculum, p. 252.
27Ibid., p. 262.
28Ibid., p. 274.
University of Missouri's guidelines, adopted in 1908, where the total of A's and B's should equal 25%, C's be 50%, and D's and E's 25% of an instructor's grades. More detailed but basically similar guidelines were adopted by the University of Iowa in 1910.

In 1915, when Rugg summarized many studies concerning the determination and distributions of grades, he noted little agreement as to what is measured by marks, whether it be (1) native ability, (2) general accomplishment resulting from ability and training, or (3) specific performance. While most experts believed in the normal distribution of ability, he found that the normal curve was unsubstantiated by objective tests. Higher grades in upper level courses, he said, were explained in part by the selective elimination of poorer students, so that ability became skewed. He analyzed some eleven proposals for grade distributions, finding that most educators accepted the normal curve as a basis for marks, but disagreed on where to draw the line to assign marks in the five division system. The general tendency, he discovered, was to place 40 to 50% of all grades in a middle group, about 20% in adjacent groups and 3-7% in the upper and lower groups, adding, however, that it was generally agreed that teachers vary widely in their standards of grading and in their reliability in marking.

From 1920 to 1930, Smith and Dobbin noted interest in individualization of instruction, in the use of standardized tests, and in the improvement of measurement techniques. The replacement of the percentage system by letters accelerated as additional evidence of the variability and unreliability of grades was reported. Since there were indications that a variety of factors in addition to student achievement affected the assigned grades, proposals were made for behaviorally-defined check lists and rating scales to replace grades and the case for a normal distribution of grades was argued.

29Ibid., p. 285.


strongly. Some of the typical literature of the time is presented below.

In 1921, Wood referred to a "psychological" test, which today is called the objective test, to determine a course grade. He called it the "new" examination, to distinguish it from the "old" examination or essay test. He accepted the unreliability of present grades, stating:

That college grades as well as high school marks are highly inaccurate and unreliable is a notorious fact upon which there is no disagreement among investigators, and which is only made more apparent as the results of each additional investigation are made public.32

He believed that better exams would improve the reliability of grades, and that the new objective exam which he proposed would provide objectivity, consistency, definiteness, comparability, and point of reference, the requirements he listed for a valid measurement scale. Grades based on a personal estimate resulting from interaction with the student, he asserted, are subjective and unreliable, as also are written examinations, which he found subjective and unreliable.

Wood also noted that some critics had suggested that all grades be simply pass and fail. He opposed this because it assumed that all inaccuracy fell in the pass portion of the range and that concealing inaccuracy by a gross classification helped little.

Several purposes have always motivated grading. On a pedagogical basis, grades were expected to be an incentive to students. On an administrative basis, they were to inform parents of the student's status; provide a fitness measure for higher education; indicate relative standing of students, so as to determine to whom credit, degrees, and honors should be given; and finally be used to judge the efficiency of instruction, the efficiency of instructors, and so on.

According to Wood's review, considerable disagreement existed about what ought to be used as bases for grading. Factors cited in the literature included effort by the student, his intelligence, his character and personality, his fitness to enter civilized society, and his improvement, if any, as well as his actual scholastic achievement. Wood believed that grades should be based on achievement alone, and that a separate record should be kept for other factors if necessary, and he insisted that there should be an accepted basis for assigning grades.

In 1925, Rogers took a look at the grading system of the University of Missouri. One significant practice there was the awarding of variable credit according to the grade,33 a practice which had been suggested by Meyer in 1915. The highest grades according to this system received greater credit, while poor but still passing grades lost a fraction of the normal credit assigned to a course.34 Rogers admitted that some faculty members abused the normal curve by giving many high grades to attract students and recommended that a student ability index be computed for each class, on the basis of extra or deficient credit which students in a present class had earned the previous term or year, and that grades be assigned accordingly. Grades might be allowed, he said, to vary somewhat from suggested distributions, particularly in small classes, but in the long run the overall norm for the classes of any instructor was to conform. This system, Rogers believed, would eliminate snap courses and student loafers who took them, and provide that honors and promotion be awarded on a fair and uniform basis. There is no record, however, that his suggestion was ever adopted.

In 1925, Kelly surveyed the grading characteristics of a number of colleges, finding that the percentage system, which assumed that the amount of knowledge a student possesses in relation to the total amount of a field is measurable, was on its way out.

A pass-fail system had been a fad in some institutions, but Kelly reported that none was still using it and


that the A to F system, though dominant, was being understood in different ways. Some college catalogs, he noted, defined letters in terms of the old percentage system, others in terms of the normal curve and the percentage of students to be assigned each letter; still others used descriptive terms as definitions, such as "exceptional" for A, "Superior" for B, and so on. Kelly believed that grades were not adequately defined, for in spite of the practice of some schools, which gave suggested distribution tables of grades to faculty, it was generally agreed by educators that a wide variation in grades existed between departments and between teachers in the same departments and even between the same sections of courses.35

It was also noted by Kelly that at some few colleges credit and quality were related, some institutions awarding more credit in courses where A or B was earned and less in the case of a D. One college simply required fewer credit hours for graduation for students who maintained a certain average, but required a certain minimum grade average as well as a specified number of credit hours for graduation and had various kinds of honors associated with high grades, and academic probation provisions with low grades.

Several difficulties associated with grading were reported by Kelly: first, students were said to concentrate on following instruction, rather than on being interested in study; second, the connection between grades and the worthwhile aims of the college was not clear; third, grades were even alleged to detract from the genuine aim of education; fourth, variation in grading standards led students to elect easy courses; fifth, the primary emphasis, as a result of interest in grades, seemed to be on memory.

Another review of grading literature was conducted in 1926 by Kerr, whose findings were not unlike Kelly's. Grading systems, Kerr found, ranged from a two-division system (pass-fail) to the 100-division system (percentile system), but in the 20's the five-division letter system predominated. Kerr reported on and approved the variation of credit according to grade level used by some colleges and argued strongly for the normal curve interpretation of letter grades, suggesting

appropriate distributions for each, based upon the figures of several other writers (3-7% A, 22-24% B, 38-50% C, 22-24% D, and 3-7% F).\textsuperscript{36}

In 1927 and 1928, Spence wrote two articles on the nature of grades. Literature on grading, he reported, was concerned primarily with evidence indicating variability between instructors, exhortations for improvement, and plans for improvement based on the normal curve.\textsuperscript{37} He believed that most often grades were assigned on a personal and private basis, with each instructor having his own rationale and justifying emotions as criteria for the assignment of a particular grade. Grades, Spence believed, should be considered relative measures, and proposed a normal curve ranking system which allowed for an adjustment of scores or grades based upon the ability of the group of students in a course.\textsuperscript{38}

Another typical article and review, written by Miller in 1928, suggested that extraneous factors such as attitude, industry, improvement, and the like be eliminated, and that grades should be based upon achievement only. Grades, he held, were very important, because they were used to assign credit, to determine eligibility, to recommend for positions. The variability of grades was also noted by Miller, and he suggested the normal curve basis for assigning them.\textsuperscript{39}

A survey of grading practices at 89 colleges in 1930, reported that three-fourths of the colleges used a five-point


\textsuperscript{38}Ralph B. Spence, "The Improvement of College Marking Systems," Teachers College Contributions to Education, No. 252 (New York: Teachers College, Columbia University, 1927), pp. vi + 90.

\textsuperscript{39}W. S. Miller, "College Marks," Problems of College Education, ed. by Earl Hudelson (Minneapolis, Minn.: University of Minnesota Press, 1928).
system and only 18 used a percentage system, the most frequent recent changes having been the switch from a percentage system to a five-point system. By this time, most colleges with the five-point system had begun to assign numerical equivalents to letter grades (A - 3, B - 2, etc.) to compute grade point averages, only a few schools holding that grade distributions should be based upon the normal curve. Actual grade distributions submitted indicated more high grades (A and B) than low grades (D and F). Six colleges gave extra credit for high grades and reduced credit for low grades. The review noted that some schools switched from one particular grading system to another at the same time other schools went from the latter system to the former.

The internal consistency of some commentators on grading at times reached the impeccable. In 1931 Davis wrote that the 6-22-44-22-6% distribution pattern for a five-point grading system was the ideal because its strict application through 12 terms (3 per academic year at his institution) would result in 48% of students graduating and 52% flunking—which happened to be the approximate number of actual graduates at his particular institution.

From 1930 to 1940, Smith and Dobbin reported that the extensive use of testing apparently overemphasized the acquisition of knowledge and paid too little heed to broader educational concerns, such as personal development; that as a result of too much testing, less tangible goals of education were being overlooked, extrinsic motivation alone was provided, and the free exchange between student and teacher was hindered through fear. Much discussion about the benefits of the normal curve, and of ways to adjust curving in various situations was noted, and many articles stressed the need to define marks specifically, to base marks on achievement only.


42Dobbin and Smith, "Marks and Marking Systems."
Crooks presented an excellent review on grading in 1933,\textsuperscript{43} noting that some educators even suggested doing away with grades altogether. Though most accepted grades as a necessity and were working to improve their use, the questioning of the reliability of marks was widespread, for confusion and inconsistency reigned regarding the basis for marking, be it achievement, effort, or other factors. Ranking was suggested by many and the resultant competition was seen as motivating students, but relative grading, based on the normal curve and a five-point scale, Crooks found was the usual recommendation. Opposition to the normal curve at that time was centered in the objection that marks had not a real meaning but only a relative meaning, and that there was no reason to expect achievement to follow a normal curve. Crooks suggested as a future direction of grading that absolute standards be developed, which neither the old percentage system or the new normal curve provided.\textsuperscript{44}

Reeves offered a fairly rigorous study to document the variability of grades,\textsuperscript{45} having surveyed practices as the University of Chicago and found considerable differences in average grades in junior and senior level courses with senior courses generally receiving higher grades. By establishing an index of student ability in each department on the basis of the GPA earned by students, he was able to show differences in standards not explainable by the attrition of poorer students or non-normal distribution of ability. Reeves states:

The lack of reliability of teachers' marks has been pointed out so frequently in educational literature as to need no further demonstration. Undergraduate students well know that the chances of getting a satisfactory grade are better in the courses of some departments and instructors than in those of others. In spite of the


\textsuperscript{44}Ibid., p. 267.

\textsuperscript{45}Floyd W. Reeves, Wesley E. Peik and John D. Russell, \textit{Instructional Problems in the University} (Chicago: University of Chicago Press, 1933).
rather general recognition of the unreliability of individual marks, averages of marks in courses have quite generally been used academically as though they were valid criteria of student progress.\textsuperscript{46}

A number of experimental colleges, following the progressive movement and the philosophy of John Dewey, were established in the 1930's. Here individual programs were established to fit each student's needs, abilities, and interest. Because students developed their own study plan with an adviser, traditional practices in regard to grades, examinations, degree criteria, and entrance requirements were de-emphasized.

From 1930 to 1957, the literature developed some new emphases according to Smith and Dobbin. For one thing, there was a growing conviction that grading practices must be consistent with educational objectives; there was less concern, as a result, about the technical-mechanical aspects of grading and more attention was given to the purposes of grading and their relation to learning. More emphasis also was given to the comprehensiveness of areas of student development which were to be included in grading with greater specificity of what was graded. More and more believed that grading should promote learning, that grades should be determined objectively and that grades should not be based on absolute standards by some teachers and on comparative standards by others. Crooks had been one of the first to express belief in absolute standards. Considerable negative criticism was given to the competitive aspects of grading according to Smith and Dobbin and finally additional evidence was marshalled in regard to extraneous factors which affect the assignment of grades, such as the likeability of the student, the sex of the student and the instructor.\textsuperscript{47}

More recent literature, organized by topical areas below, presents very few new problems: pass-fail grading, grading based on absolute standards, even abolishing grades altogether are all suggestions made before. Terminology has changed indeed, and emphases have shifted, but the basic issues for the most part are the same.

\textsuperscript{46}Ibid., p. 48.

\textsuperscript{47}Ann Z. Smith and John E. Dobbin, "Marks and Marking Systems."
Grading Systems: Actual and Proposed

In this section an attempt is made to present and analyze various actual and hypothetical grading systems, some of these have been noted above in the history of grading. Literature relating to these grading systems and research results is also at times included.

Grading Systems in Use--Surveys

A 1934 survey reported on 88 colleges. Ninety-four percent used letter symbols, and most of these had a five-letter system with four passing grades, and the remaining six percent used numbers based on 100. The letter symbols were sometimes undefined (12 schools), frequently defined by single-word descriptions (57), sometimes defined by numbers (22), and infrequently defined by the normal curve (2). A survey of 34 institutions in 1943 found the same general practices with some few changes: only one school used a percentage system; one-sixth used pluses and minuses with letter symbols. The actual percentage of letter grades assigned, however, varied widely at different institutions: A's - 3.5% to 22.5% and F's - 1.5% to 16%.

In the 1960's, two surveys reported some new practices. The first, which surveyed 54 institutions (mostly Eastern and prestigious) cannot be considered representative. About one-fifth of them used pluses and minuses; three used special symbols such as H (honors), HP (high pass), P (pass), and two reported using no grades except for retention and probation. The majority employed the traditional five letters with associated numerical quality points. The reviewer saw chaos in the systems used, was troubled about

---


49 Ibid., p. 263.

50 Winona M. Perry, "Are Grades and Grading Systems Comparable From One Institution to Another?" Journal of American Association of College Registrars, XVIII (1943), pp. 159-165.
the variation in practice, and suggested that re-examination was needed.51 The second survey reported a longitudinal study of 300 representative institutions and compared 1957 practices with those of 1967, noting that the primary change was the adoption of provisions for pass-fail or pass-no credit by one-sixth of the institutions.52

Three surveys have been reported on in the 1970's. The survey of 549 institutions in 1970 showed that most used traditional letter grades and grade point averages, and that about half had adopted some form of the pass-fail option as part of their overall system. Some few institutions recorded only satisfactory grades (i.e., A, B, C, no record), used all pass-fail grading, or substituted written evaluations for letter or number grades.53 In 1971 a study of 435 institutions revealed that about three-fourths of all those schools used some sort of pass-fail grading while a meager three-and-a half percent used total pass-fail grading. Some of the latter, however, had added "Pass with High Honor" and "Pass with Honor" to restore some recognition of quality work.54

The largest survey, including a study of 1301 member institutions, was conducted by AACRAO, and published in 1971, with 52% reporting completely traditional grading systems and 46% combining traditional with non-traditional. A few institutions had adopted "non-punitive" grading and did not report F on the transcript; 96% did report F. At 54% of the institutions, however, a student was allowed to repeat a course, in which case the original grade was replaced by the repeated grade. Registrars indicated that grading changes


were accelerating at their institutions and predicted that grading systems as a result will become less traditional.55

In general, the surveys indicated the total abandonment of the percentage system, the overwhelming predominance of the five-point, A to F system. Any number of variations, as might be expected, were found in the basically traditional system. The study showed that the majority of schools surveyed include some pass-fail grading option, that a few use plus and minus, and that some have begun to adopt the practice of not recording grades below C or D. Almost all institutions compute an overall grade point average by assigning a numerical equivalent of letter grades. While there is some indication that an increasing number of variations on the traditional system is being tried, these are still the exception. Only a handful of institutions have reported a completely non-traditional grading system.

**Traditional A, B, C, D, F System**

As the surveys above have indicated, this system is at present dominant in American higher education, being the standard, in contrast to which other systems are called non-traditional. Three levels of satisfactory pass are incorporated into the system, together with one level of unsatisfactory pass, and failure for which no credit is granted. GPA is computed by assigning numerical equivalents to letter grades and is used for a variety of internal purposes (such as dismissal, graduation, honors) and external purposes (selection by graduate schools and employers). Faculty members are generally free to assign grades as they please, providing they do not give all students in a class either high grades or low grades.56


56Buchman, "Grading Systems Across the Country ..."
Grading System Based on Normal Distribution or Curve

Grading on a curve results from assigning or adjusting a student's grade according to his relative achievement, relative, that is, to the achievement of other students in a course. Grading according to a normal distribution has the additional implication of adjusting grades to coincide with the statistical curve of normal probability. A normal curve for grading is based upon the assumption 1) that factors such as academic aptitude and motivation, and other factors that result in or cause what is measured in a grade, are distributed normally within the student population that constitutes a class, and 2) that the conditions for learning, such as length of study time, course materials, instructor contribution, are uniform. If, for example, a less able student were allotted additional time, greater achievement might result, and the assumption of normal distribution would not hold.

From the history of grading it is clear that the assumptions about the desirability of a normal distribution of grades were strongly endorsed by many, even if infrequently followed closely in practice in regard to the percentage of grades which various writers suggested be given at each level. However, several writers have stated their opposition to this form of grading for a variety of reasons. Considerable evidence, they say, has been gathered to indicate that the assumption of a normal distribution of ability in each class of students is not valid, and several surveys

57 DeWitt J. Davis, "... Normal Curve System on Failures and Grade Values."
Kerr, "Grading Systems."
W. S. Miller, "College Marks."
Rugg, "Teachers Marks and Marking Systems."

Crooks, "Marks and Marking Systems: A Digest."
Kerr, "Grading Systems."
have presented strong empirical evidence to support this.\textsuperscript{59} As a result of his study of the question, Vandervelde recommended that "The standard of distribution of grades should be based upon the quality of student graded by a given grader."\textsuperscript{60}

Though faculty members on the whole may not grade on the basis of a normal distribution as technically defined, it has been shown definitely that they do grade according to their own implicit or explicit curves. Several studies explored the relationship between increasing ability in successive entering freshman classes at a college and the average freshman GPA which was assigned. All studies showed essentially the same thing, that despite rising ability GPA remained the same.\textsuperscript{61}

Another study vividly demonstrated the cumulative effects of grading on a strict curve. In the second term of

\begin{itemize}
\item \textsuperscript{59} J. H. Greene and C. R. Hicks, "Do College Class Grades Follow a Normal Distribution?"\textit{College and University,} XXXVI (Spring, 1961), pp. 296-302.
\item R. H. Riffenburgh, "Fitting Asymmetric Student Grade Distributions,"\textit{Journal of Educational Research,} LIII (Dec., 1959), pp. 123-129.
\item Arvo E. Juola, "Illustrative Problems in College-Level Grading,"\textit{Personal and Guidance Journal,} XLVII (Sept., 1968), pp. 29-33.
\item \textsuperscript{60} Vandervelde, "A Neglected Consideration in Grading Systems," p. 317.
\item \textsuperscript{61} Kenneth M. Wilson, "Increased Selectivity and Institutional Grading Standards,"\textit{College and University,} XLVI (Fall, 1970), p. 51.
\item Juola, "Illustrative Problems in College-Level Grading."
\end{itemize}
a two-term course, continuing students who had earned C's or higher had their second-term grades adjusted downward, in many cases even to D and F, to take up the slack in lower grades which had been created by the departure of the poorer students. A table showing these results is presented in Table 1 of Appendix C. The conclusion of these investigations is simply that the faculty as a whole had no implicit standards of good and poor performance.

Several problems are suggested by the results, and the question arises, should grades reflect achievement relative to class performance or relative to independently defined standards of quality? Also, to the extent that grades are relative to class norms, are there educational consequences (such as competitiveness, dysfunctional self-criticism, anxiety) when students are unable to reconcile perceived quality and amount of work with the grade received? Juola believed that grading practices were chaotic, with wide variations among sections, instructors, and departments without justification. In order to correct the situation he believed three steps were necessary: (1) studies to find out what is actually happening; (2) development of a philosophy of grading which is conveyed to all instructors; (3) systematic safeguards to implement decisions derived from the philosophy to apply to diverse grading problems. He stated:

To continue the present laissez-faire view toward grading is unrealistic. Instructors must have information about college-wide and department-wide grading practices and methods for coping with unique situations. Otherwise, grades and grading distributions must reflect the idiosyncracies of individual instructors no matter how serious and sincere their intentions may be.

In summary, the practice of "grading on a curve" is an established fact, the basis for any particular curve generally being left to the individual instructor. While they

62Juola, "Illustrative Problems in College-Level Grading."


64Juola, "Illustrative Problems in College-Level Grading," p. 33.
raised some question about the desirability of normative grading under any circumstances, the primary interest of these writers is that the "individual" curve of each instructor conform with a rational institutional policy. In the next section one such rational institutional policy is considered.

Grade Distributions Adjusted for Student Ability

A number of writers who believe that grades should be assigned in relation to the achievement of other students nevertheless suggest that some adjustment should be made to take into account differing ability levels among groups of students. As early as 1925 as well as later, there were proposals to make such a correction, on the basis of student average grades in other classes, of IQ, or of aptitude/ability test scores.  

Another approach which involved a highly mathematical procedure for achieving desired results, used standard deviations related to grades, deriving the correction factor on the basis of aptitude entrance tests. Grade distributions, according to this procedure, would be suggested to the instructor, and the range of grades assigned would be expected to be consistent with the suggested distribution.

One writer actually computed this correction factor, using the SAT and high school rank. His results showed great

65 Rogers, "A Uniform Grading System."
Spence, "The Improvement of College Marking Systems."

discrepancies between actual and assigned grade distribution ranges as well as the magnitude of adjustment required. His findings are presented in Table 2 of Appendix C.

No information was cited in the literature which indicated that any institution actually practiced a system of grading whereby distributions or curves were formally adjusted according to ability.

Grading by Student Rank

Another variation of normative grading or grading on a curve involves ranking of students. In 1915, it was suggested that an instructor report only student rank in a course. The registrar would change the rank to percentiles for transcript purposes. Fifty years later the same system was suggested by Seawall with some additional refinements: GPA would be the average of all percentiles in courses taken and graduation standards would specify some minimum average percentile, i.e., 25 or higher. Seawall was criticized by Carlsson, since class sections are frequently not typical cross-sections by ability and ranks would not be accurate. Surprisingly, Seawall admitted the inaccuracies of letter grades as charged but was unconcerned since he believed that the inaccuracies would cancel each other.

Similar systems were also suggested by other persons, but again the literature does not record any such system actually having been used.


67 Meyer, "The Administration of College Grading."

68 Allan Carlsson, Reply to F. Seawall, "Quality Grading System," College and University, XLIV (Fall, 1968), pp. 13-16.

69 Ebel, Measuring Educational Achievement.

Contract System of Grading

Several writers have suggested a grading system based on contracts with students. Briefly, the instructor and student agree on specific objectives to be achieved usually according to some plan established by the instructor. A specified minimum is required for D, additional work for C, a paper in addition, perhaps, for B, and so on. Each student will fulfill different requirements according to the "contract."71

Advantages of such a system include: clear specification of what is expected by the instructor; emphasis upon individual success; democratic choice; non-competitive environment; stronger motivation for students. Problems encountered were: minimal motivation in students contracting for low grade; multiplicity of assignments becoming burdensome; tendency for quantity of work to substitute for quality of work; lower achievement as measured by common exam with students in a traditional course; rigidity flowing from specifications.72 In one trial reported on it seemed apparent that this system affected the interaction of student and instructor in creating feelings of mutual respect and trust.73 Another investigator found these results to follow from the procedure: (1) students' growing personal involvement in the establishment of objectives; (2) less coverage of material, but greater depth; (3) increased personal responsibility by students; (4) greater usage of traditional and non-traditional


Robert W. Taylor and others, "Effects of Contingent Grading on Student Course Work" (Bloomington, Ind.: Indiana University, 1969), ED 049 671.

learning resources; (5) sharing by students in evaluation of their achievement; (6) heightened motivation, creativity, student-teacher rapport; (7) much greater demand on the instructor's time.74

On an individual basis, instructors are able to use this approach within a traditional grading system of their institution. Contract grading, however, has not been used on a systematic basis by an entire institution, nor has it been extensively evaluated except on a personal opinion or anecdotal basis.

Obviously, the contract system does imply a completely different approach from grading on a curve or normal distribution, for in general, the distribution of contract grades may take any form, depending upon how many students complete contracts at various levels. It should be noted, too, that contract grading not only involves new methods of grading, i.e., non-competitive and individual standards, but also a new approach to teaching, which relates to each student in an individual way and greatly reduces reliance on uniform and standard procedures, such as the lecture method.

Multiple Dimension Grading

One strong objection to traditional grading has been the one-dimensional nature of a letter grade, which many writers cited below have pointed out. As a result, various schemes have been proposed for reporting additional information. One writer who attacked the problem argued for eight dimensions to be rated on a pass-fail basis, where students must achieve pass on the dimensions appropriate to the course. Dimensions he suggested included recall, comprehension, application of knowledge, evaluation, analysis, synthesis, attitudes, and psychomotor skills.75 Another writer suggested three grades for each course, based upon a student's effort,


ability, and performance. Elbow recommended an eight-factor system, with three levels of ranking; namely, weak, medium, and strong. Elbow's factors were command of course information, understanding of central ideas, imaginative and creative use of subject matter, verbal strategy, thought strategy, class contribution, growth over the semester, and diligence and effort. He stipulated that the faculty should report on those factors deemed important to the course.

A last writer, perhaps with tongue in cheek, suggested a special form of multiple-dimension grading, namely, that two grades be recorded for each course, one submitted by the instructor and the other by the student. This was seen as an interim measure until grading was abolished. The use by any institution of this grading system is not reported in the literature.

Descriptive Grading

Descriptive grading is a somewhat grander or expanded approach than multiple-dimension grading. Proponents of this approach would abolish letter grades, ranks, percentiles, or similar marks and in their place use a brief prose description of the student's achievement in a course. Wolff, a radical critic of traditional grading, wanted each student to receive detailed criticism but no grade, rank, GPA, or anything equivalent. Some writers suggested combining the description with a basic pass-fail or credit and no-credit

---

76 Max S. Marshall, "Triangular Grading," College and University, XLIII (1968), pp. 143-149.


Marshall favored reporting "salient characteristics and outstanding features" of student achievement, suggesting that a "Q" for questionable be assigned to students failing to do satisfactory work, along with a descriptive explanation of whatever problems existed.81

The proponents of these systems saw them as reducing undesirable competition and anxiety and at the same time providing students and other users of grades information much more valid for use both in learning and in selection functions, since the student's transcript would supply the instructor's descriptions, perhaps from several sentences to a half page for each course, and corresponding notations of either credit or no credit granted. There would be no ranking or grade point averages on the transcript.

At least five institutions have had their systems of descriptive grading described in the recent literature of grading. At one college a comment system was adopted in which students were evaluated on the basis of stated goals of a course, rather than on a single scale relative to others in the class. In addition to their written comments, faculty members used a symbol to indicate whether or not credit was


earned. At another institution the instructor, student, and the student's counselor all wrote reports dealing with the student's achievement; these written reports were placed into a cumulative record maintained in the registrar's office. The registrar abstracted from the total record of information about various courses which then became the transcript of record for the student. One can only speculate as to how many persons were employed in this registrar's office.

At Brown University, descriptive grading was added to a letter-grade system of A, B, C, No Credit or S (Satisfactory), No Credit. Instructors submitted a detailed report of the student's performance in the field of concentration, including a description of both the program and the student's performance. This detailed report was made part of the student's permanent record and GPA was not computed. At a fourth institution, instructors' written evaluations took the place of numerical grades for juniors and seniors in some advanced courses.

Finally, the University of California at Santa Cruz employed a university-wide pass-fail system for all courses, coupled with written evaluations of the student's performance. (Letter grades were allowed by way of exception in some courses with permission) Written evaluations ranged from one-word euphemisms substituted for the usual

---

82 Margaret A. Faust, "In Opposition to Conventional Grades," (Claremont, Calif.: Scripps College, Feb., 1971). ED 051 753.


84 Milton E. Noble, "Advantages and Disadvantages Associated With the "No-Fail" Grading Systems," College and University, XLVI (Summer, 1971), pp. 717-726.


86 Howard Shontz, "The Grading System at University of California, Santa Cruz--a Critique," College and University, VL (Summer, 1970), pp. 494-501.
letter-grades to several paragraphs of descriptive prose. The typical undergraduate transcript ranged from one to ten pages, depending upon the number of courses taken at the university and the characteristics of the student's instructors. According to the report of the Academic Senate, students under this system tended to focus more on the material of a course and its specific personal importance, than on an extrinsic grade reward. Improved morale and attitudes and reduced cheating were reported; instructors were more oriented to students as individuals and academic standards were apparently maintained.87

In spite of such results, only a few institutions have adopted descriptive grading, perhaps because of the practical problems arising from the handling of so much information, and from the added burden put on instructors in formulating such evaluations. In discussing scales and measurement for such a system, Flanagan noted:

While very comprehensive descriptions can be defended on theoretical grounds, they sometimes become too cumbersome and complicated for practical use. A practical reason for adopting a simplified descriptive procedure in reporting test behavior is to provide a set of symbols which may be readily manipulated and understood. A long verbal record of responses to a series of test situations may be completely descriptive, but it is often quite difficult to communicate, compare, or combine with other observations.88

This is precisely what proponents of descriptive grading think ought to be avoided.

Self-Grading

The literature records some attempts by individual instructors to allow students to assign their own course

87Frank Andrews, "Report on the Grading System at UCSC--a Critique" (Santa Cruz, Calif.: University of California, Committee on Educational Policy, 1970). (Mimeographed.)
grade. In three studies, students suggested their final grade, but the grade of the instructor was the one that counted. Results of these studies are inconclusive. One study indicated that superior students tended to under-rate and below average students to overrate themselves, but another reported that students in these categories tended to estimate their final grade more accurately. One study concluded that self-evaluation was satisfactory and a better educational experience, while others concluded that students cannot satisfactorily grade themselves. A study by Rose indicated that student grade expectations, both near the beginning as well as toward the end of the semester, were much higher than term-end grades.

In some experiments, when instructors actually submitted the student's grade as the course grade, the student-assigned grades were invariably considerably higher than those which would have been assigned by the instructor. Noting this, one writer concluded that self-grading would reduce the validity of grades for external selection functions, but would enhance the process of learning. A second writer noted that some students achieved more and some less under this plan, and that achievement was sensed more personally by students, and a third believed that any additional


learning achieved in this process was not worth the cost. He used the same special evaluation methods in the same course the following year and, assigning his grades, found even better educational outcomes.\textsuperscript{93}

Evidently the results of self-grading are mixed, and reports regarding it cannot be considered reliable, since so few studies were made and under such widely differing conditions. Generally self-grading when it is adopted is combined with much more comprehensive attempts at evaluation and feedback to students on their achievement, a fact which may by itself account for most of the positive outcomes recorded.

**Pass-Fail Grading**

A very large amount of grading literature has been concerned with pass-fail grading, moreso than with any other topic in recent years. From about 1960, that system was the most popular and practically the only significant change in grading practices. Its popularity blossoming in the early sixties, by the early seventies some form of pass-fail grading was used by the majority of all institutions, though at the same time they still employed basically traditional grading systems.

**General Characteristics of Pass-Fail Grading**

Four general surveys of pass-fail grading were reported in the literature from 1967 to 1971.\textsuperscript{94} All indicate that the


pass-fail option is restricted by such a variety of regulations that the dominance of traditional grading is left basically intact. General restrictions include: excluding freshmen; permitting only one P/F course per term; setting a maximum limitation on credits earned under P/F; excluding courses in the student's major field; restricting such grading to elective courses only; restricting it to students in good standing. Some institutions require permission of the instructor for the innovation, and the instructor may exclude some courses entirely from P/F.

While a grade of P was not counted in the student's grade-point-average, about forty percent of all institutions included an F received under the P/F option. A variety of registration procedures were used, with most institutions requiring the student to choose between the P/F or letter-grade options within the first weeks of the term or even at registration. About one-third of the institutions involved did not indicate to the instructor which students were registered on a P/F basis, presumably so that the evaluation will be more objective. About one-third used P to include A to C while others included D as passing.

A myriad of variations used by participating institutions were reported in the literature, one being the use of...


"AACRAO Survey of Grading Policies in Member Institutions."


G. H. Massey and others, "Transcript Adequacy: P's, F's and Other Letters," College and University, XLIV (Summer, 1969), pp. 520-527.


Kirker Smith, "Pass-Fail: Were Your Hypotheses Valid?" College and University, XLVI (Summer, 1971), pp. 549-556.
of letter grades internally, while reporting all courses taken as either passed or failed to external users.\textsuperscript{96} Another departure employed a total P/F system for freshmen and traditional grading for all other students.\textsuperscript{97}

In general, the pass-fail grading option has become very popular as a modification of traditional grading systems. With few exceptions, however, it was introduced, as we have seen, with a variety of restrictions and limitations such that the dominant position of traditional letter grading and the accompanying grade-point-average was not threatened.

Objectives of Pass-Fail Grading

Diverse reasons have been cited for introduction of the pass-fail option. These varied from institution to institution, but the more general reasons offered, cited in the approximate order of their frequency, include: encouragement of students to explore a new or unfamiliar field without concern for an adverse grade; allowing the student to learn without pressure, competition, or emotional strain; encouragement of intrinsic motivation and intellectual


curiosity. 98 Two writers noted the freedom this option allowed the student to control the allocation of study time or to produce unevenly rather than to excel in all areas since he might choose the area in which to be outstanding and opt simply to pass in other areas. 99 This rationale seems to be closely related to the concept of discrete and varying abilities and that one need not achieve in every area. Other outcomes were that such grading might minimize the fear of failing and counteract some of the undesirable effects of the current grading system. 100 Similar reasons were forwarded in other articles on the topic. 101

98 Brick and McGrath, Innovation in Liberal Arts Colleges, p. 77.
Ericksen, "Grading Evaluation."

Ericksen, "Grading Evaluation."

100 Quann, "Pass-Fail Grading: What Are the Trends?" Claude Simpson and Charles Quann, "Pass-Fail--What Are the Trends?" College and University, VL (Summer, 1970), pp. 484-494.

When students have been asked to report their actual purposes for taking courses on a P/F basis, the four most frequently cited reasons were: (1) to reduce the competitive pressures of letter grading (61% responded above average); (2) to complete a general university requirement where interest or background was lacking (56%); (3) to remedy a situation in which there is insufficient time to study for the traditional letter grade (44%); (4) to explore outside the major without jeopardizing the grade-point-average (42%).

Another survey indicated that lack of time (70%), interest in exploring a new area (44%), or insufficient background (23%) were primary reasons for choosing P/F. Others included lack of interest in a required course, desire for general knowledge without details, and pursuit of a special interest without grade worry.

One original thinker argued for the adoption of P/F only in well-developed, specific fields such as science and math, where one could absolutely determine whether a student achieved mastery and should receive a "pass." A-F grading was seen as appropriate for less precise fields, where more gradations of judgment and certainty are called for.

---

Brick and McGrath, Innovation in Liberal Arts Colleges.

Ericksen, "Grading Evaluation," p. 5


Primary reasons cited by institutions in adopting the pass-fail grading option centered about encouraging students to explore new and unfamiliar areas without worry about receiving an unsatisfactory grade, and at the same time somehow improving the general learning situation. The latter goal is expressed in vague terms, with little specification as to how it should come about.

The reasons already listed as to why students choose pass-fail overlap, but obviously differ considerably in order and emphasis from the officially stated reasons.

Objections to Pass-Fail Grading

A number of objections have been raised against the pass-fail option. These include: reduced student motivation; problems in such matters as the Dean's List, academic probation, and suspension; the computing of grade-point averages; admission to graduate study; differential effects on grade-point average.106 Disadvantages to the student, pointed out by one author, include the lack of feedback on progress; the lack of information to guide the student's placement in future learning activities and to assist him in being placed in a suitable kind of work; and the absence of real evidence to show pass-fail grading does achieve its stated objectives for the student.107

Some opponents of traditional grading also objected to pass-fail grading, contending that it is not clear that P/F significantly alters student competition and anxiety, or greatly affects teaching.108 It may result, they claimed,


108 Charles Pascal, "Methods of Grading and Models of Teaching" (McGill University, Center for Learning and Development, 1970). (Mimeographed.)
in poorer evaluation by the instructor, since fewer distinctions are made,\textsuperscript{109} and can imply acceptance of almost any work by the student, or can, on the contrary, result in the application of a most extreme negative decision, since only two unmodified opinions are allowed for final judgment.\textsuperscript{110}

The concern about P/F arises then from two distinct sources. In the context of traditional grading and its purposes, P/F injects uncertainty and unreliability. From the non-traditional viewpoint, P/F fails to resolve adequately the problems of grading. A limited P/F system either does not basically change the older system or, from another perspective, it is even more restrictive than traditional grading in that it reports less information about student achievement. Ericksen, a supporter of P/F, described three areas of concern: social issues, including such implications as ranking for selective service; administrative objections, which relate to P/F's effects upon the use of grades for transfer, selection, academic honors, and the prediction of future success; learning issues, including effects on teaching, student motivation, and effects on creativity.\textsuperscript{111}

**Outcomes of Pass-Fail Grading**

It is patent that evaluation of the use and outcomes of pass-fail grading has been extensive; unfortunately, however, much that has been said is subjective, uneven or spotty in coverage, and at times inconsistent. Nevertheless, some convergence of conclusions emerges from the array of reports which is consistent with the results of the more carefully conducted studies.


\textsuperscript{110}Marshall, \textit{Teaching Without Grades}.

\textsuperscript{111}Stanford C. Ericksen, "Pass-Fail Grading," \textit{Memo to the Faculty}, No. 22 (Ann Arbor, Mich.: Center for Research on Learning and Teaching, University of Michigan, 1967).
Exploration of New Fields of Study

Encouragement of students to explore new fields of study is the primary reason offered for the adoption of pass-fail grading, but the literature almost without exception indicates a complete failure to achieve this purpose. Although one survey of student opinion showed that about half of all students interviewed had elected not to take certain courses, from concern about a lower grade-point-average which might follow, yet introduction of the P/F option seemed to change this very little.

Furthermore, one rather extensive study showed that the amount of exploration outside of a student's own subject areas is about the same for students who use the P/F options as those who do not, and a series of briefer studies consistently reported that very few students use P/F to broaden their knowledge and intellectual experience by taking courses not required of them. At one institution, only 7% of 292 students who had used the P/F option indicated that they would not have taken the course if the option had been unavailable. At a second institution, the corresponding percent

112 Suslow, "Pass-Fail Grading At Berkeley . . .," p. 60.


George L. Melville and Eleanor Stamm, "The Pass-Fail System and the Change in the Accounting of Grades on Comprehensive Examinations at Knox College" (Galesburg, Ill.: Knox College, Aug., 1967).


Kirker Smith, "Pass-Fail: Were Your Hypotheses Valid?"


Keith Wharton, "P-N and Non P-N Students: Reasons for Electing the Option; Graduate School Plans" (Bureau of Institutional Research, University of Minnesota, Aug., 1969).

114 Cromer, "Investigation of Student Attitudes Toward Pass-Fail Grading System."
was 19%, and at a third college, though some 47% of students registered on a P/F basis said they were taking the course because of interest, they were not asked if they would have taken the course without this option.116

In summary, it can be concluded that students by and large do not use the P/F option to explore new areas of study which they would otherwise not have taken out of concern for lowering their grade-point-average.

Relief from Pressure, Anxiety, Competition

Many of the objections to traditional letter-grading related to the anxiety, competition, pressures, and similar effects experienced by students. The P/F option, by specifically avoiding indications of gradations of quality, was presumed to minimize or reduce these undesired side effects.

Two studies assumed that if P/F reduced anxiety, then students using this option would be those with higher anxiety levels, but while one of the studies reported that students with higher grade-point-averages had higher anxiety levels, neither was able to show that high test anxious students chose P/F more frequently.117

Other studies have been based for the most part on surveys of student opinion concerning the perception of anxiety and pressure felt under the P/F option. All such studies reported that students consistently judge that there is less pressure and anxiety in courses taken under the


116Stallings, Wolff and Machr, "Test Anxiety, Grade Utilities, and the Pass-Fail Option."

117Ibid., p. 8
pass-fail option. Several studies also reported that a major (if not the major) reason given for choosing the P/F option was to reduce pressure and anxiety. Having more time (by taking a course under P/F) is related to experiencing less pressure, and this reason for the choice was also advanced by students. One study of faculty opinion indicated that the faculty also believed that the option reduced the competitiveness of students somewhat and moderately relieved pressures; some instructors added that students were more relaxed in class.


Wharton, "P-N and Non P-N Students."
Cromer, "Investigation of Student Attitudes Toward the Pass-Fail Grading System," p. 3.
Suslow, "Pass-Fail Grading At Berkeley . . .," p. 45.


Suslow, "Pass-Fail Grading At Berkeley . . .," p. 45.


121Suslow, "Pass-Fail Grading At Berkeley . . .," pp. 42-43.
Although most reports are based on subjective opinion, the general consistency of all studies, both large and small and from a variety of institutions, substantiates the conclusion that the P/F option does reduce competitiveness, tension, and pressures which arise from grading and time requirements.

Satisfaction of Individual Needs

Not many studies have been concerned with this particular matter, but those that have suggest: (1) that students have greater latitude to fulfill felt personal needs, rather than to compete with other students;\textsuperscript{122} (2) that they learn "different" things, i.e., concepts, for example, while disregarding details;\textsuperscript{123} (3) that generally they can use their own judgment under the freedom of P/F to tailor their own work.\textsuperscript{124}

Proponents of P/F maintain that this latitude in allowing the students to work according to their own felt needs rather than following the rigid schedule of activities expected by the instructor is very desirable. Indeed, in one study the writer found that students were neglecting laboratory skills and theory, and that only problem-solving seemed to remain a student concern.\textsuperscript{125} Opponents of P/F agreed that various course objectives can be safely ignored by students, especially if D is included in the range of P. However, they argued that this is not desirable unless it is assumed that these established course objectives are dysfunctional or of no educational importance to the P/F students.

Motivation and Achievement

One reason considered as important in providing the P/F option has been to provide more suitable, meaningful

\textsuperscript{122}Saunders, "A Student's Perception of the Non-Traditional Grading Pattern."

\textsuperscript{123}Pascal, "Pass-Fail Grading . . ."

\textsuperscript{124}Strong, "Pass-Fail Grading System."

\textsuperscript{125}Ibid.
motivation for students. While motivation, time commitment, and achievement are not identical, they are related. Numerous studies in this area, especially regarding P/F and achievement, have been conducted, some based upon self-reporting by students and faculty, many others arising from a study of grades made at schools where faculty members are not informed about a student's P/F status but must assign regular grades, which later are converted to P/F by the Registrar. Thus comparisons can be made between P/F and non-P/F courses.

In the opinion surveys, students opting for P/F admitted to greater absenteeism, less participation, completion of less work recommended by the instructor, and spending less time than they otherwise would have committed. In the opinion surveys, students opting for P/F admitted to greater absenteeism, less participation, completion of less work recommended by the instructor, and spending less time than they otherwise would have committed.126 Studies of faculty opinion reported the same consistent findings, that students enrolled in their courses under the P/F option do not work to capacity.127

---

126 Cromer, "Investigation of Student Attitudes Toward the Pass-Fail Grading System," p. 3.
Lindenwood Colleges, The Academic Program (Self Study) (St. Charles, Missouri: Lindenwood Colleges, 1969-1970.)
Wharton, "P-N and Non-P-N Students."
Bain, Hales and Rand, "Does Pass-Fail Encourage Exploration?", p. 18.

Ibid., p. 17.
Five studies are reported where the instructor was unaware of students' P/F status, and P/F students were compared to non-P/F students, while no controls for student ability were introduced. One large study found that P/F students tended to do better; a second could see no differences between the two groups; two smaller studies concluded just the opposite. The fifth, involving large numbers of students and over a hundred t-test comparisons, found that P/F students do distinctly poorer work, percents of grades for P/F and non-P/F students respectively in the arts and sciences being computed as follows: A - 3.8% and 21.3%; B - 19.7% and 31.5%; C - 44.6% and 25.5%; D - 20.2% and 5.3%; and F - 5.7% and 2.4%. On a four-point scale the GPA scores determined by this survey for P/F and non-P/F students are 1.83 and 2.36 respectively.

The discrepancy between these five studies probably resulted from non-random choices of student subjects and from lack of any control for estimating student ability. Evidence indicating that ability levels actually are higher among P/F than among non-P/F students could explain the results of the first two studies and make the interpretation of the last three studies even more significant.

Other studies have established some control for the ability level of the student by comparing the grade of the P/F student in the P/F course with those in non-P/F courses. In these experiments also the instructor was unaware of the P/F status of students. All of these studies show clearly that the student achieved less under the P/F

---


129 Kirker Smith, "Pass-Fail: Were Your Hypotheses Valid?"

130 Robert E. Thayer, "Do Low Grades Cause College Students to Give Up?" (Long Beach Calif.: California State College, April, 1971). ED 054 725.

Melville and Stamm, "The Pass-Fail System and the Change in the Accounting of Grades on Comprehensive Examinations at Knox College."
option, and that the difference can vary anywhere from two- to six-tenths of a grade point.131

A review of the work of some 900 students controlled in a different manner, by previous GPA, ACT scores, course load, etc., likewise demonstrated that students received significantly lower grades under the P/F option, and the author recommended as a consequence that any subject based on cumulative learning should not be offered on a P/F basis with a D passing level.132 A more experimental study allowed some students to take all or one course on a P/F basis and compared them to controls who wanted, but were not allowed, the P/F option. P/F students received significantly lower grades (before conversion to P/F), 1.67 as compared to 2.26 on a four-point scale. Even after returning to conventional grading, the P/F students continued to receive lower grades and no compensatory improvement was noted in grades they received in other courses.133


Melville and Stam, "The Pass-Fail System and the Change in the Accounting of Grades on Comprehensive Examinations at Knox College."


In addition to these overall findings, some evidence had been established to indicate that specific provisions of the P/F option also affect student achievement. In one case it was reported that inclusion of the D grade under F, rather than under P, resulted in better performance on the part of students, who thereby have the incentive to do at least C work.\textsuperscript{134}

The weight of all studies, both those comparing P/F to non-P/F students and those controlling for ability, indicates that achievement levels are significantly lower under the P/F option as measured by the grades assigned. Further, other comparisons such as those on absenteeism, workloads, assignments completed, and student opinion point to the same conclusion. If it is intended that pass-fail grading enhance traditional academic achievement as measured by any of the above factors, it must be considered a complete failure.

Other Outcomes

One outcome of pass-fail grading seemed to be its effect on student grade-point averages. While one study cited above indicated that students using the P/F option achieved normal grades in their regularly-graded courses, most studies suggested that students were able to raise their grades in these courses. At Wellesley, for example, the most striking result of P/F was the rise in GPA of many students who were apparently able to eliminate their potentially lower grades.\textsuperscript{135} It was also noted above that students chose the P/F option for reasons associated with their inadequate preparation and lack of sufficient study time. If students by and large take courses for P/F that they would have taken anyway without the option, and if their reasons for the choice as cited above


\textsuperscript{135}Cromer, "Investigation of Student Attitudes Toward the Pass-Fail Grading System," p. 5.
are accurate, then it is correct to assume that students are selectively eliminating low grades so that GPA will rise. At least two other studies, in addition to that made at Wellesley, suggested that this is the case.  

Another reported P/F result was the benefit simply resulting from a change, any change, from the traditional system, since it is hypothesized that the mere act of grade reform has a positive effect on student morale. Other writers, it was seen, denied this, arguing that P/F grading is simply a token handed to mollify them while it preserved the traditional system pretty much intact.

Closer Student-faculty relationships and improved student morale have also been suggested as P/F outcomes. At the University of California at Santa Cruz, which adopted pass-fail grading for all courses combined with written evaluations by instructors, less cheating, more humane education, improved student morale, closer student-faculty relationships, and other such factors were reported. At the California Institute of Technology, freshmen took all courses on a P/F basis, and lower attrition, higher morale, and better performance at the sophomore level were reported. Situations there are not comparable to the usual limited P/F option and so more may have been achieved at these institutions. However, it remains difficult to evaluate these and other similar claims made for the usual P/F since such claims are primarily personal opinions concerning


137Pascal, "Pass-Fail Grading . . .," p. 12.

138Saunders, "A Student's Perception of the Non-Traditional Grading Pattern."


140Strong, "Pass-Fail Grading System."
poorly-defined and fairly amorphous qualities that have not
been empirically verified. Further, such worthwhile results
may occur not as a regular and expected result of the P/F
option, but rather as a transient side-effect of the intro­
duction of a novel artifact.

There may be some tendency among students using the
P/F option to register for heavier course loads than those
carried by students who do not. This is not well explored,
but one study found that P/F students carried significantly
heavier course loads;141 this would be consistent with the
finding that students plan to and actually do spend less
time on the P/F course.

Summary of Pass-Fail Outcomes

Undoubtedly, pass-fail has resulted in some changes
within the grading environment; in its usually limited form,
however, the overall effect seems to amount to no more than
a gesture toward students to prove the flexibility and will­
ingness of institutions to respond to expressed student con­
cerns. In fact, it has allowed traditional grading with all
of its implications to remain more or less unchanged.

A review of the literature on pass-fail grading was
completed by Quann and reported in 1974. He called the move­
ment an "unsuccess story," since it failed in its most
widely-expressed goals, namely to encourage students to ex­
plore new areas without fear for their GPA, and to reduce
anxiety and pressure stemming from competition for grades.
Quann maintained that neither goal has been substantiated
and that, if anything, the results may be negative, since
students do less work in P/F courses.142

Acceptance of Pass-Fail by Graduate Schools

The usual criteria for acceptance into graduate school
have been undergraduate grades and achievement and/or aptitude

141 Stallings, Wolff and Machr, "Test Anxiety, Grade

142 James C. Quann, "Pass-Fail Grading--An Unsuccess
Story," College and University, IL (Spring, 1974), pp. 230-235.
tests. An important concern for institutions adopting the pass-fail option was the effect it may have upon the acceptance of their graduates for advanced study. As a result, a series of surveys of graduate deans and of persons responsible for admission into graduate or professional schools have been conducted. Without exception, those interviewed indicated that with a few pass-fail grades the students' opportunities are not greatly affected, but that more than a few such grades will negatively affect their admissions possibilities and also their chances for financial aid.143

There was some indication that a significant number of non-traditional grades on a student's transcript forced schools to rely more on other bases for admissions decisions,


"AACRAO Survey of Grading Policies in Member Institutions."


using test scores and recommendations, for instance.\(^{144}\) Graduate school staffs reported reluctance, however, to base an admission decision on test information alone, and judged undergraduate GPA to be the best indicator of future success.\(^{145}\) The consensus is that P/F grades in electives primarily, and in general requirements secondarily, created a lesser problem than would arise in attempting to interpret such grades in the student's major.\(^{146}\)

Studies in this area are very consistent in their findings. For better or worse, they agree, a student is likely to experience difficulty in obtaining admission and financial aid at graduate and professional schools in direct proportion to the number of non-traditional grades appearing on the transcript, since graduate and professional schools definitely prefer letter grades in evaluating and making decisions about prospective students. They concede that

---

\(^{144}\) William W. Hassler, "Results of Pass-Fail Questionnaire Sent to Graduate School Deans" (Indiana, Pennsylvania: Indiana University of Pennsylvania, 1969). ED 037 168 MF.


\(^{145}\) E. F. Tragesser and others, "Are New Developments in Achievement Testing an Adequate Answer to Pass-Fail Grading Systems?" College and University, XLIII (Summer, 1968), pp. 565-567.


students with only a few non-traditional grades, generally less than 10%, and with letter grades only in their major fields, do not experience great difficulty.

One institution using a total pass-fail system, but supplementing the pass grade with written evaluations for each course, reported that 56% of the graduates said they had no difficulty getting into graduate school; 35% thought they may have been affected adversely, and 9% responded affirmatively. It should be noted that a student has no direct way of knowing whether his admission was adversely affected, unless such specific information was received by him from the graduate school.

Student Attitudes Toward Pass-Fail

In spite of the variety of P/F options offered them and the differences in questions asked them in P/F studies, students indicated that they have reached a clear consensus to the overall desirability of the P/F option, being overwhelmingly in favor. Some are satisfied with the present nature of the limited option, and others would like to expand it to include all general requirements, but researchers prove that pass-fail grading has become a very popular innovation with students.

Faculty Attitudes Toward Pass-Fail

Research concerning the general popularity of P/F with faculty members has not been as extensive as that made

147 Andrews, "Report on the Grading System at UCSC."

148 Cromer, "Investigation of Student Attitudes Toward the Pass-Fail Grading System," p. 4.

Stallings and Smock, "Pass-Fail Grading Option at State University."

Morishima and Micek, "Pass-Fail Evaluation."


150 Quann, "The Pass-Fail Option: Analysis of an Experiment in Grading," p. 11.
with students. Only two studies have taken faculty opinion as a primary focus, and they noted that reaction is mixed, with some grudging acceptance of the limited form of P/F, and at the same time with some concern about how students use P/F in practice.151

Summary

As the pass-fail option was being introduced on a wide-scale basis in 1968, Stallings and others specified several areas of research which were needed to determine its desirability. These areas included effects on student achievement; on attendance, attitudes, and outside preparation for class; on courses elected for pass-fail.152 While the literature on pass-fail grading may benefit from additional experimental research in some instances, the volumes of literature consistently indicate that valid answers have been obtained. The primary objective of pass-fail, to encourage broader exploration of new fields by students, has by and large not been achieved, student motivation, effort, and achievement are lower; less pressure is experienced by students in such courses, and time saved there is spent not on their P/F work but on traditionally-graded classes or in taking extra courses; students are freer to pursue their own purposes, as opposed to instructor-established objectives.

Judging by the studies done thus far, faculty opinion toward pass-fail is mixed; student opinion is very favorable. They indicate, too, that limited pass-fail grades are not a great problem in obtaining acceptance to graduate school or in seeking employment.

It probably should not be a surprise that pass-fail in its limited form has had questionable results. Pascal

151Suslow, "Pass-Fail Grading At Berkeley . . .," p. 40.
Ibid., pp. 92-93.
Stallings and Smock, "Pass-Fail Grading Option at a State University."

suggested pass-incomplete in all courses in the context of programmed instruction which would also allow students to participate in the setting of objectives.\textsuperscript{153} His solution may not be desirable for various reasons, but there is no question but that such comprehensive approaches would at least have a much greater impact than "traditional" pass-fail.

**Honors, Pass, Fail**

A three (or sometimes four) level grading system suggested by some writers is a modification of traditional grading systems in the direction of pass-fail grading. Angus, for example, suggested Honors, Pass, and Fail categories, arguing that instructors cannot make accurate distinctions between five categories, i.e., whether a borderline student should receive B or C. Another writer suggested three categories called "Mastery of a subject, Pass or Competent in a Subject, and Attempted," in the context of a mastery approach without time constraints.\textsuperscript{154} In his extensive review of the pass-fail option, Suslow concluded there should be three categories—"Outstanding, Successful, and Not Acceptable." The highest category would include the top 10 to 15% of students and the lowest category would generally include students formerly receiving D or F.\textsuperscript{155} All of these systems were devised to bring about more uniform grading practices and more validity and fairness in grading for students.

A few institutions actually adopted such systems. Yale used Honors, High Pass, Pass, and Fail;\textsuperscript{156} Whitman

---

\textsuperscript{153}Charles E. Pascal, "Alternatives to Traditional Grading Procedures: Research Findings and Implications" (Montreal: McGill University, Center for Learning and Development, 1969).

\textsuperscript{154}W. B. Cameron, "Deliberate Radicalness," *Journal of Higher Education*, XXXII (March, 1961), pp. 152-156.

\textsuperscript{155}Suslow, "Pass-Fail Grading At Berkeley," p. 5.

College used Highest Honors (HH), Honors (H), and Passing (P). Whitman's HH and H were slightly more difficult to earn than A or B had been, and P included the former C and half the D range. In both instances, difficulties in communication with other schools and fear of possible discrimination against their graduates led to a reinstatement of traditional grading.\textsuperscript{157} Westminster College adopted Distinction (DN), High Pass (HP), Pass (P), and No Credit, with DN and HP more difficult to achieve than A and B had been; No Credit included the former D and F. GPA was not computed but students were ranked.\textsuperscript{158}

Such grading systems, using terms instead of the A to F terminology, are essentially similar to traditional grading, for though the five traditional levels are reduced to three or four, meaning and context remain as provided by traditional grading. At the same time, some options introduced in these systems, such as not recording unsatisfactory work and allowing for repetition of work, do of course depart from traditional practice. GPA and ranking may or may not be computed when such systems are used, but the computation is certainly possible.

Non-Punitive Grading

"Non-punitive grading" is the name used by proponents of the growing practice of not recording F and sometimes also D on the student's permanent academic record, several surveys in the late 1960's showing that some institutions were then practicing aspects of non-punitive grading.\textsuperscript{159} In 1971, the extensive AACRAO survey of grading practices indicated that 4% of institutions surveyed did not record F on the transcript, and that 54% replaced an original


\textsuperscript{159}L. McHugh, "Repeat Course Policy," \textit{College and University}, XLIV (Spring, 1969), pp. 279-282.
unsatisfactory grade with the repeated grade. Non-punitive grading was second in popularity only to pass-fail grading as a modification of traditional grading, though concern about grade inflation has recently slowed the trend. Individual institutional modifications of this approach include NC or No Credit or No Credit or No Credit or No Credit or No Credit or No Credit for unsatisfactory or incomplete work, with one institution allowing withdrawal from a course at any time. Some institutions reported the No Credit grade on the transcript while others do not, but it is never included in the student's GPA.

Arguments for a non-punitive grade were diverse. First, it eliminates a double penalty, in that an F not only results in loss of credit, tuition, and effort, but must also be balanced by one A or two B's in order to achieve the generally required GPA of C. Second, elimination of F may reduce the anxiety which can hinder learning and

160AACRAO Survey of Grading Policies in Member Institutions.


162Noble, "Advantages and Disadvantages Associated With the 'No-Fail' Grading Systems."

163Brick and McGrath, Innovation in Liberal Arts Colleges.


166"University of Minnesota to Drop Use of Failing Grades."

167Noble, "Advantages and Disadvantages Associated With the 'No-Fail' Grading Systems."

168Lawrence G. Smith, "Non-Punitive Grading."
discourage exploration of new fields. 166 In addition, Erickson believed that this sort of system will occasion a profound reorientation of the education process, in requiring more careful establishment of criteria for success, in relating the teacher and students more closely in the conditions for learning, and in reducing the gap between evaluation and grading. He also noted that, according to learning theory, punishment serves a limited and even negative learning function. 167

Other writers stress the system's non-competitive approach to learning; 168 the reduction of number of students on probation; the orientation toward rewarding appropriate behavior rather than toward punishment; 169 and the increase in grading standards resultant on not granting credit for D, thus giving students a better chance to earn C or above. 170 Commenting on non-punitive grading, one author dubbed traditional grading "a relic of the Protestant ethic that all who fail should be punished." 171

167 Erickson, "Grading Evaluation," p. 5.
170 Douglas L. Fels, "Let Each Become All He Is Capable of Becoming/Let's Stop Certifying Failure," What's Going on in United Methodist Junior Colleges, No. 6 and 7 (May and Sept., 1972).
Traditional grading by implication suggests the notion of "negative" knowledge. In order to meet the generally required C average for good standing and graduation a student receiving a D must compensate by showing "surplus" knowledge, so to speak, in another course by obtaining a B, and so on. The student receiving a D or F makes "negative" progress toward graduation. However, it is unlikely that the student learned absolutely nothing, much less that he lost knowledge, even though admittedly he achieved less than course objectives. Fels attacked this interesting dilemma of traditional grading.172

A second paradox follows from traditional grading, that of a marginal student capable of only C work who is currently achieving at that level but who has earned numbers of D's and F's in early college work. He should be advised to transfer to another institution, since any number of C's would never quite raise the student's GPA to 2.00, but since D's and F's do not transfer, the student would be able to graduate at the second institution on the basis of the same work which would not satisfy graduation requirements of the home institution.

Arguments against non-punitive grading include these: that without F there is a loss of motivation; that it is unrealistic, since failure is a part of life; that it makes inefficient use of resources by failing to eliminate poor students in favor of more qualified students.173 Concern was also express about the loss of the "veracity" of the transcript when poor achievement is not reported.

These objections are not validated by research, and some cannot be, since they are essentially philosophical. The objection concerning motivation is unconvincing since the motivation of good grades remains in this system and students must still be concerned with loss of credit, tuition, and time even though the poor grade is not permanently recorded. The argument that education should be modeled after real life and include competition and failure was discussed earlier. In contradiction to the supposition that

172Fels, "Let Each Become All He Is Capable of Becoming/Let's Stop Certifying Failure."

173Lawrence G. Smith, "Non-Punitive Grading."
fear of failure leads to success, there is evidence to indicate that success in learning generates more success, and vice versa. Institution resources can still be used efficiently under non-punitive grading by a provision calling for dismissal of a student after he fails to achieve credit in a specified number of courses.

Difficulty concerning "veracity" of the transcript arises from the fact that the "consumer" of a student's transcript may be led to believe that all work, satisfactory and unsatisfactory, is noted, but a clear explanation of non-punitive grading accompanying a transcript would prevent this. While all information desired by the consumer may not be provided, proponents of non-punitive grading maintain that the legitimate interest of the consumer should only extend to the competencies as achieved by the student. In light of the tenuous relationship between grades and any criterion measure of success (such as work) other than further study, the absence of this information would seem to make little difference.

Non-punitive grading, second only to the limited pass-fail option as the usual departure from traditional grading, has a number of implications not always clearly formulated. It implies, for example, some specifications of educational objectives for grading, as opposed to relative grading or grading on a curve, and it is a less competitive model and is oriented to a successful academic experience by students. As of this moment it has generated less literature and research than the limited pass-fail system, though its implications are potentially much more far-reaching. Understandably, as the practice of non-punitive grading spreads, this situation may change.

Elimination of Grades

In addition to the many forms and variations of grading systems already discussed, some writers have suggested seriously the option of doing away with grades. In reviewing the literature of grading as early as 1933, Crooks noted that even then some critics would completely do away with grades. By 1971, at the annual American Association of

174Crooks, "Marks and Marking Systems: A Digest."
Higher Education Conference, some 129 of a total of 572 respondents indicated that they did not believe any kind of grading or ranking of students was necessary.\textsuperscript{175} A summary of a conference on grading systems reported the general wish that grades could be eliminated, although the thought was not made as a serious proposal.\textsuperscript{176}

Several respected critics, in fact, have made such a proposal. In 1962, Cummins suggested dropping all grading or allowing the student to grade himself, with the instructor simply providing feedback to the student.\textsuperscript{177} Axelrod noted the widespread dissatisfaction with the grading system; listing the negative effects he believed grading had on the instructional process, he concluded: "... American college and university education in general, and the education of creative young people in particular, would be greatly benefitted if the standard grading system were eliminated or drastically modified."\textsuperscript{178} Later, Blaylock blamed some of the student unrest of the sixties on the grading system and argued for the elimination of grading or at least for a deemphasis of grading until its elimination could be achieved.\textsuperscript{179} Around the same time McMahon furthered the trend by asserting that grading, or characterizing a student as a C- or D-student, is dehumanizing, and that grading should be dropped.\textsuperscript{180}

\textsuperscript{175}College and University Bulletin, XXIII (May 1, 1971), p. 2.


\textsuperscript{179}Mabry G. Blaylock, "Student Unrest From a Middle Ground," Improving College and University Teaching, XIX (Summer, 1971), pp. 211-213.

\textsuperscript{180}Michael B. McMahon, "Positivism and the Public Schools," Phi Delta Kappan, LI (June, 1970), pp. 515-517.
Wolff, in *The Ideal of the University*, developed his position against grading in greater detail by defining three levels or kinds of "grading": criticism—a desirable form of feedback on performance; evaluation—measure of performance against an objective standard of excellence by which a comparison can be summarized as yes or no or by finer gradations; and ranking—the relative comparison of the performances of students to determine an ordering of comparative excellence. He considered "criticism" to be at the heart of education, but held that evaluation and ranking are extraneous to education and have been adopted for purposes extrinsic to education, such as selection for employment or graduate study. In conclusion, he advocated a three-year undergraduate program, designed with its own integrity and without grades of any kind.\(^\text{181}\)

What approximates a no-grade system was suggested by Hyman. Grades, he suggested, should be given only in departmental work, and on only two occasions: the first would be on an exam designed to qualify a student to enter upper level work, and the second would be given on a final exam at the end of a student's program, to determine what had been achieved. In Hyman's system, no comparative grade would be given in any individual course.\(^\text{182}\) Becker, Geer, and Hughes suggested giving no grades to the very able students, who did not need the motivation provided by grades, and some form of non-published grades to poor students, who would be overwhelmed by poor grades.\(^\text{183}\)

The advocates of no grading have for the most part failed to develop a careful rationale for an educational program without any form of grades, nor have they dealt with the problem of quality and standards or studied the difficulties arising from the need to satisfy demands of employers and graduate schools. Regrettably, such very real practical problems have received instead no adequate study, but were


merely dismissed, despite their importance to educators and students alike.

Individualized Grading Systems

Some few recent writers have formulated a suggestion which is unlikely to stir the enthusiasm of many registrars, and that is to employ two or more grading systems which would be compatible with the needs of different students or with the changing needs of the same students at various times. Haverford in 1967 adopted a system of using regular grades for freshmen and sophomores for internal purposes only, i.e., when they are reported to the student, adviser, and dean. Official transcripts, it was reported, should show only a list of courses taken, with a notation if the student has failed or withdrawn. In the junior and senior year, it was agreed, grades should be recorded as usual. The theory is that the first two years were a transition time in which the student was encouraged to explore the curriculum freely, a time in which it was important to minimize grade consciousness and anxiety. The system changed in the upper years, when the student was thought to be more ready, and when grades were assumed to have additional significance as devices for evaluation of the student by graduate and professional schools.184

Other Grading Systems

Obviously, a great number of grading systems are possible and a fair number have been tried. Combinations of distinct grading systems, which can multiply the total possibilities, are also mentioned in the literature of grading.

The variable credit system noted above was proposed in the early Twentieth Century. As early as 1915, Meyer had recommended granting greater credit for higher quality student achievement.185 In 1972, Sheleff, however, suggested


an updated variable credit plan, proposing that one credit be awarded in a course for proficiency on a midterm and final exam; two for the above, plus a short report; and three, four, or five for specified additional work, but there would be no grading, other than the awarding of credit. In some courses, Sheleff conceded, it might be appropriate to require two or three credits as a minimum to be earned; for example, in courses taken in a major. Such an approach, the author believed, would eliminate the complicated grading system, provide better motivation than P/F, and allow greater freedom and flexibility for the student. Unanswered questions are whether substitution of more courses by poorer students is a satisfactory substitute for quality, whether students would want to complete a whole course for only one credit, whether an additional short report for an extra credit is in any sense equal in quantity to what is required for earning the first credit, and so on.

Hyman suggested that the tyranny of grades could be avoided while maintaining academic standards by using all P/F in the classroom, while assigning letter grades in departmental exams not related to specific courses as such to students intending to major or minor in a field. Those not majoring in any area would receive all P’s and get a "Pass" degree.

Purposes or Objectives of Grading

Logically any discussion ought to begin with the purposes or objectives of grading. For the complex matter of college grading, however, it is necessary to provide some prior perspective by examining the history of grading and various grading practices. The objectives of grading are, of course, crucial to the entire grading discussion. Yet what objectives are to be served by grading is not a matter of simple or even of complex empirical investigation and justification. Rather the question of objectives is, to a


187 Hyman, "Grades and Academic Standards."
very large extent, a matter of prudential judgment, since some objectives may and do conflict with other objectives.

The enumeration of grading objectives has not changed greatly in fifty years. Wood, in 1921, described both pedagogical objectives, which provided motivation to the student, and administrative objectives. From the student's viewpoint, his list included provision of feedback to parents, a scholastic fitness measure, and a basis for credit, degrees, and honors. Among objectives for the administration, he cited supplying data for vocational guidance as well as a basis for judging the efficiency of instruction and of instructors.188 Vandervelde, in 1937, described as primary objectives supplying information to student and parents and to the institution regarding credits earned and quality of work; as secondary objectives he listed the procuring of an indicator of a student's fitness for further work, feedback to the student, motivation for the student, and a source of information for employers.189

More recent writers generally describe the same set of objectives. Raimi outlined three categories: (1) to the student, feedback for learning; (2) for the institution, feedback to the instructor on the learning process; (3) for the external public, information desired by graduate schools, employers, and parents.190 Other writers in similar lists of objectives mentioned grading as contributing to effective and efficient instruction; supplying feedback information and motivation for students; meeting institutional purposes (such as the assigning credit, the fulfilling of graduation requirements, and the writing of recommendations); and aiding in guidance and selection functions.191


Even Nathan Glazer, a strong critic of grading and of the educational system, conceded that grades served some purposes, such as selection for graduate school and work on a merit rather than a status basis. He stated: "... if grades did not exist, they would have to be invented because they perform a function which is not easy to dispense with, and on the whole they perform the function well enough, and economically." Erickson, on the contrary, did not believe that grades assisted the teaching or learning function since their major function, he said, has been to provide information to sources removed from the classroom, such as graduate schools and employers.

Ebel discussed the importance and the nature of the problem of grading and in doing so considered a number of issues, including the need for grades. He stated:

The uses made of marks are numerous and crucial. They are used to report a student's educational status to him, to his future teachers, and to his prospective employers. They provide an important basis for crucial decisions concerning his educational plans and his occupational career. College education is expensive, and there are not enough opportunities for college students to go around. If we are to make the best possible use of our educational resources, and of our student talent, it is essential that each student's educational progress be watched carefully, and reported as accurately as possible. Grade reports serve somewhat the same function in college education that financial statements serve in business. In either case, if the

---


193 Ericksen, "Grading Evaluation."
reports are inaccurate or unavailable, the venture is likely to be inefficient, and ultimately to fail. Grades also provide an important means for stimulating, directing, and rewarding the educational efforts of students.\textsuperscript{194}

According to Ebel, therefore, grades had reporting, selection, and motivating functions. His comparison of grades to a financial statement, though unacceptable to some grading proponents, suggests a variety of interesting comparisons.

The most extended discussion of the purposes or objectives of grading and underlying implications was developed in an article by Feldmesser called "The Positive Functions of Grades."\textsuperscript{195} Here grading objectives were divided into first-order functions, which provide unique and useful information for students, and second-order functions, which were derived from some other functional need. In the first sense, the grade as "summative evaluation" attempted a wholeness and synthesis of many separate judgments and summarized the multidimensional evaluation made by the instructor during the course. All things considered, such a grade indicated to the student whether he did well or poorly and showed him whether this course or field, as compared to others, is productive for him. The same logic was applied by Feldmesser to the cumulative grade point average which he says indicated to the student how he was doing on the whole and, all things considered, whether continued study was appropriate.

Each kind of grading, according to this authority, served different purposes. Second-order functions, he proposed, are twofold. First, the obligation to provide a fair grade which can be defended put an obligation upon the instructor to carry out a careful plan of evaluation in each course, an activity that an instructor might otherwise seriously neglect, although Feldmesser saw this evaluation process as essential to a sound educational program.

\textsuperscript{194}Robert L. Ebel, "Basic Considerations in Grading the Achievement of College Students," (Lansing, Mich.: Michigan State University, Sept., 1966), p. 2. (Mimeographed.)

Second, such central reporting or grading, which can have specific desirable or undesirable consequences for a student, induced or motivated the student to be responsive to the objectives and goals of the course as formulated by the instructor. Feldmesser did not believe that a student was completely competent to decide what he will and will not study in a course or what was or was not relevant to a field of study, and the experiences with P/F cited above corroborate his contention that students may to some extent feel freer to disregard the course plan without traditional grading. Finally, Feldmesser pointed out that in order not to spend unreasonable effort on the reporting function, some easily handled system, such as the traditional system, is required.

Obviously the objectives sought from grades have an important impact on the forms grades take, on grading practices, and on the uses to which grades are actually put. As a result, the incompatibility of various objectives needs to be considered. For example, to what extent do selection objectives and learning objectives clash? And if they do, what is the nature of the priorities to be observed? Literature and research relating specifically to the goals and purposes of grades are unfortunately very meager, but almost every general article on grading in some sense refers to objectives, usually in the fashion of an undefended and selective laundry list offered as a springboard to promote certain points of view. Various grading objectives cited by numbers of articles are: extrinsic uses--1) in a student's selection for graduate school, 2) in his selection for employment, 3) as a reporting function for parents; institutional uses--1) as a basis for awarding credit, 2) as a basis for awarding honors, 3) as an aid in determining progress and continuance in school, 4) as a measure of efficiency of instruction, 5) as an indication of student's preparedness for advanced work in an area; teaching/learning uses--1) as motivation for the student, 2) as feedback to student on progress, 3) as feedback to instructor on students' progress, 4) as a basis for guidance for the student, relative to future study plans, 5) as preparation of the student for life in a competitive system.

An important issue in the consideration of grading objectives dealt with implicitly and in different ways by

---

196 Ibid.
the writers cited above is just what is meant by grading. If grading is simply defined as the process whereby a symbol, be it a letter, number, or word, is finally assigned to a student's work in the form of summative evaluation, then some of the goals listed above are partially inappropriate. If, on the other hand, grading represents this summative evaluation process as well as an account of the exams, papers, and intermediate evaluations which are included in the "grade," then the objectives of grading become more comprehensive in scope. This distinction is discussed more thoroughly below, under evaluation, but for the moment the broader concept of grading is used.

In this context a summary list of grading objectives which are used in this study is presented:

1. Feedback, teaching tool for instructor.
2. Feedback, learning tool for student.
3. Motivation for student.
4. Means of rewarding behavior desired by instructor (other than academic achievement—such as attendance, attitudes, etc.).
5. Mark of recognition for competencies or knowledge achieved.
6. Selection/screening tool for college and graduate school. Basis for selection, awarding honors, retention in college, etc.
7. Selection tool for employer.

Determinants of Grades

While the proposed objectives of grades are numerous and perhaps conflicting, the final assessment of any objectives will be affected in practice by their reliability and validity, and in accord with the bases on which they are assigned. This latter concern is discussed in the present section. In this context the determinants may be primary, the immediate basis for assigning a grade, or secondary, contributing to the primary determinant. This topic, like most others concerning grading, is not new. In the history of grading it was noted that moral judgments as well as other factors consistently influenced grades. A review of grading practices by Crooks in 1933 showed that a variety of factors in addition to achievement affected grading,
although he stated his belief that only achievement should be considered.  

Academic Ability

Academic ability obviously is expected to and does strongly affect achievement as measured or recorded by grades. In fact, this relationship is by far the most thoroughly studied issue in the area of grading. As the results are well known, the literature is not reviewed here. In general, two indicators of ability are measured, high school rank and some ability test, such as the ACT or SAT. High school rank generally correlates with freshmen GPA around .70 and the ACT or SAT around .50. Adding the ACT or SAT results to high school rank by means of a multiple regression formal increases the relationship to GPA only slightly. The reason for this is that the global nature of the high school rank already includes most of what is measured by ACT or SAT in addition to many other factors, such as motivation and even the ability to make friends and influence instructors, or possibly to cheat well. Presumably all those factors leading to higher grades at the high school level are also generally operable at the college level.

While a correlation on the order of .70 is significant and several qualifications are important, such a relationship explains only half of the causal relation associated with the college GPA. The ability portion of the high school rank as measured by ACT or SAT explains only one quarter of the causal factors. Further, the correlation between high school rank and test scores decreases beyond the freshman year, so that the relationship between these measures is reduced further.

A unique study by Boldt in 1970, in which he attempted to examine the factorial bases of grades, was conducted at


two graduate schools of business using ATGSB verbal and quantitative scores and grades obtained in various courses. Boldt's analysis indicated that a one-ability theory explained the data as well as those assuming two or three abilities. The author concluded that grades were factorially simple and only one underlying ability need be postulated. It must be noted, however, that the sample of students and type of courses were highly selective, and the findings may apply only to such similar cases.

Other Abilities

It is customary to think of ability as a factor which is general in nature, but is probably more correct to assume several abilities exist which are only loosely correlated with one another. Both the ACT and SAT measure several areas, including the traditional verbal and mathematical abilities. The relatively modest correlation between ability and GPA, especially after the freshman year, may be due in part to the varying and specific abilities required. Hoyt, for example, believed there were many types of student growth and development which were relatively independent of each other.

In this area, Richard, Holland, Lutz, and James undertook extensive research which was reported in a series of articles. Basically, twelve scales developed to


measure "non-academic" achievement, included: leadership achievement, social participation, artistic achievement, social service achievement, scientific achievement, business achievement, humanistic-cultural achievement, religious service, musical achievement, writing achievement, social science achievement, and speech and dramatic achievement. The conclusions of these researchers, based upon a large sampling, were that non-academic accomplishment can be assessed with moderate reliability, that both academic and non-academic accomplishment can be predicted to a useful degree, and that non-academic accomplishment is generally independent of academic potential or achievement.²⁰² A recommendation follows the research results:

Further, the results imply a need for a broader, or different, definition of both the nature of human talent and the nature of higher education. There are many kinds of human accomplishment, and each kind is likely to benefit from some type of higher education, although not necessarily a highly academic type. In other words, our results imply a need for a wide variety of colleges, many, if not most of them relatively unselective except on dimensions clearly relevant to their particular emphasis. Measures of academic and non-academic accomplishment would then be used in helping students find an

---


appropriate college, rather than being used in selecting
students for a single college.203

Creativity

In a study of National Merit finalists, Holland
examined the relationships between three criteria of aca-
demic and creative performance and a variety of other vari-
ables. He found that creative performance at the high school
level occurs more often among students who are independent,
intellectual, expressive, asocial and consciously original,
and who have high aspirations for future achievement. The
relationship between academic aptitude and creative per-
formance, Holland decided, was negligible.204

Axelrod in his book, The Creative College Student:
An Unmet Challenge, argued strongly that traditional grading
rewarded the non-creative student for standard sorts of
work, contending that grades and creativity either have no
relationship or are negatively related.205 Pemberton also
found some evidence to indicate that creative students re-
ceive lower grades.206

These judgments should be interpreted in a general
sense, for though there obviously are courses in the area of
art, writing, and other fields where a student without cre-
avtivity would be severely handicapped, it may still be true,
however, that creative and individualistic work is not
highly rewarded in most situations.

203 Ibid., pp. 353-354.

204 John L. Holland, "Creative and Academic Performance
Among Talented Adolescents," Journal of Educational Psychol-
ogy, LII (June, 1961), p. 146.

205 Axelrod, "The Creative Student."

206 W. A. Pemberton, "The Grade Point Average: Snark
or Boojum?" (Neward, Delaware: Student Counseling Service,
Student-Faculty Interaction

Only a few research efforts have been directed as to how student-faculty interaction might affect GPA. A study by Runkel in 1956 indicated that students whose attitudes were colinear with the instructor did earn higher grades, and when Hamlish, in 1954, studied the effect of similarities of teacher-student personalities on grades, he found that students receiving high marks tended to describe themselves as the teacher described them and, conversely, they were described by the teacher in terms similar to the teacher's self-description. The teacher's ratings of high-ranking students were more like the students' self-descriptions than were such ratings of low-ranking students. Hamlish concluded that personal interaction can affect grading, where interaction is possible in a course. Lewis, in a study of factors used in grading, found that no instructor consistently used the same factors from student to student, and suggested that a student's personal characteristics affected the grading practices of every instructor. In his review of research on the prediction of academic achievement, Lavin stressed the instructor-student interaction:

A student's grade is more than something that characterizes him as does his score on a personality inventory or an intelligence test; that is, it is not simply a personal characteristic or a trait. Rather, a grade should be viewed as a function of the interaction between student and teacher. In short, it is one index of this social relationship. So considered, it is clear that if we want to predict a grade, we must know something not only about the student (his


ability, his personality, his values), but about the teacher as well.210

Other variables described below could be listed here since student-faculty interaction is involved.

Faculty Criteria

What does a faculty member formally consider in the evaluation of a student in a course and the assignment of a grade? While "student achievement" is the general criterion, what is actually included under that term, and whether or what other factors than strict achievement ought to be included, is hotly debated.

In a general analysis of grading, Erickson indicated that what went into a grade depended upon the teacher. Some grades included penalties for weak character, absences, tardiness, overdue assignments, and similar factors. Others penalized or favored the slow starter. Some faculty attempted to grade each student in terms of his individual progress, i.e., the rate of progress was given more weight than the final level of achievement.211 Buxton suggested that some instructors graded low to keep students in line and to assure high standards and scholastic character. Others graded "easy" to make friends with students. Still others may reward particular aspects of student work with high grades.212

The primary problem of present grading, according to Elbow, was that any grade given in a course can be considered too ambiguous, being based, as it is, on many factors which were not explained and which varied from professor to professor and between upper and lower level courses. To justify his claims, he presented the following lengthy list of factors which he believed could consciously or unconsciously

---


211 Erickson, "Grading Evaluation," p. 2.

determine a grade: command of course information; memory; understanding of the central concepts of the course; logical/conceptual intelligence; application of the central concepts of the course to new instances, seeing new implications; creativity, imagination, intuitive insight; effectiveness of verbal strategy; effectiveness of thought strategy; curiosity; permanence of learning; integration of course matter with what is already known; growth or improvement; utilization of potentiality; potentiality for further development; judgment; diligence, effort; moral trustworthiness; likeableness; enjoyment of learning.213

Ebel, as one of two basic shortcomings of grades, cited the lack of clearly defined, generally accepted, scrupulously observed definitions of what grades should mean,214 commenting: "One of the important requirements of a good marking system is that the marks indicate as accurately as possible the extent to which the student has achieved the objectives of instruction in the particular course of study."215 Various indicators or measures may be used, he argued, as long as they are related to achievement.

In a second article Ebel noted that too often a grade was determined by the student's pleasantness, class participation, oral and writing skills, successful apple polishing, or improvement. In disagreement with the common practice, he insisted: "Grades should not aim to report nor be greatly influenced by the student's attitude, his effort, or the amount of improvement he has made. They should simply report the level of achievement he has reached."216 Changed attitude, effort, etc., according to Ebel, should be reflected in the grade only when they are specifically related to the instructional objectives of the course. Others reported the same variety of criteria being used and added more to the list, among them the use of term papers, class participation, type of tests used, varying relative weights

214 Ebel, Measuring Educational Achievement, p. 401.
216 Robert L. Ebel, "Basic Considerations in Grading the Achievement of College Students" (Lansing, Mich.: Michigan State University, Sept., 1966), p. 9. (Mimeographed.)
of factors, methods chosen to observe accomplishment, accuracy of testing instruments, and so on.217

A few empirical studies have been completed in the area of grading criteria. At one institution the grading criterion was the level of achievement of course goals at the time the student took the final examination, but at other institutions the end-of-semester achievement was considered in the light of other factors, such as working to one's potential and improvement during the semester, and allowances were often made where financial, physical, or psychological conditions in the student's life, known to the instructor, prevented him from doing his best.218 In a freshman English course surveyed for one such study, it was found that teachers whose class grades correlated well with an ability measure emphasized test grades, outside assignments, a knowledge of grammatical rules, and honesty, while those teachers whose grades had a low correlation based grades more on neatness, class attendance, information feedback, and class attitude.219 At Harcum Junior College a survey indicated that in addition to academic achievement a large number of faculty assigned some grade-weight (as high as 70%) for class participation. Other factors cited as influencing assigned grades included student interest, cooperation, attendance, attitude, outside work, extra-credit work, enthusiasm.220

At Michigan, although the majority of faculty cited achievement of course objectives as the basis for grades, yet a variety of other factors were also listed, including


amount of progress, quantity of work done, acceptability of
the student for advanced work, fraction of required work com-
pleted satisfactorily, quality of work done, knowledge of
course material, skill in using knowledge, meeting course re-
quirements, intellectual honesty and integrity, and position
in distribution of test scores. It was also shown that fac-
culty in various fields often placed differing emphases on
the several definitions listed as achievement-related qual-
ities. While these results were interpreted as strong
evidence of the lack of consistent meaning, it should be
pointed out that many of the meanings can be rather closely
related to the dominant meaning, i.e., fulfillment of course
objectives.

A second part of the study at the University of Mich-
igan was of more interest in this regard. The same 50 fac-
culty were asked to check factors which they believed to
affect grades. The factors and number of checks were:
ability to apply logical criticism (34); effort in course
(26); skill in writing (25); amount of participation in
class discussion (21); interest in course (20); attitude to-
ward course (15); promptness in turning in written work (15);
ability in verbal expression (15); regularity in attendance
(13); attentiveness in class (13); mental ability, indepen-
dent of achievement (10); regularity of preparation for
class (9); tardiness (4); ability to get along well with
people (4); none of the above (7). The authors noted that
many items were checked only by those who believed that
grades should measure other things than just achievement,
since some factors implied that marks may be considered as
reward or punishment for good or bad behavior, and they
stressed the need for the precise definition of the in-
tended meaning of grades.

The matter of grading criteria, or the factors on
which a grade ought to be based, is a subject that merits
much more research than it has received to date. This is a
matter distinct from other considerations, such as standards
and other factors which may intentionally or unintentionally
affect a grade. The actual degree to which inconsistent

221Robert M. W. Travers and Norman E. Gronlund, "Mean-

222Ibid., p. 372.
criteria affect assigned grades has not been determined, but from the information available it seems clear that grading criteria are not uniform. Some factors cited above, such as the standard of comparison, are not basically substantive criteria but rather technical issues, which are discussed below.

Personality Traits of Students

A relatively new area of study, most of which has been conducted since 1960, has examined the effects of personality traits upon grades. Holland, one of the first researchers in the area, reported in 1960 that nonintellectual variables, such as superego, persistence, and deferred gratification, can be used to predict the academic achiever, and that students with certain personality variables might do better at one school than at another, depending upon the characteristics of the institution. A year later he reported that students who were persevering, sociable and responsible, and those whose parents were somewhat authoritarian, were more frequently academic achievers.

A series of studies has linked conformity as well as other variables to grades. One study indicated that students chosen by faculty as achieving institutional goals and as likely to be successful were somewhat more conforming and routine noncreative persons. It also showed that a good number of characteristics considered desirable by faculty were unrelated to grades. The CPI measured traits, including as categories achievement by conformity and achievement by independence. A provocative study classified courses as those which rewarded conforming behavior, i.e., highly-structured courses, and those which rewarded independent behavior, i.e., idea-oriented, less structured, and more

---


individualized courses. This work found that students high on achievement by conformity did better in the first course, and those high on achievement by independence did better in the latter type of course.  

Two studies compared the personality traits of students in multiple-section courses who tended to do well on in-term instructor evaluations, versus students who did better on a common end-term examination. The higher instructor-grade group, the survey determined, were more compulsive, conforming, rigid, and insecure than the second group, which tended to be flexible and adaptive. The instructor-oriented group, it was thought, made an effort to impress their instructors and thereby could influence what was to a large extent a subjective evaluation.

Dogmatism, or open-mindedness and close-mindedness, and grades was the subject pursued by Conway, who used Rokeach's Dogmatism Scale and a scale of life goals. He found that open and closed-minded students taught by open and closed-minded faculty received the same grades; likewise, students with the same or different goals as the faculty received the same grades. However, closed-minded students in low agreement with the goals of their teachers received higher grades. The writer speculates that such students participate or verbalize more actively in class to protect a

---


threatened belief system and that such participation earned a higher grade for them.228

In 1968, Korn was able to distinguish between three groups of students on the basis of what the curriculum meant to them. Some were basically oriented to getting good grades, others saw the curriculum in terms of its being useful for their careers, and the third group had an intellectual interest. Showing that each group had different personality characteristics, he concluded that there were significant differences in the ways that students learn and in their underlying patterns of motivation for learning.229

An important series of studies of personality traits, conducted at the University of Delaware, was reported in 1970.230 One of these indicated that students getting higher grades than expected on the basis of test scores were relatively more dependent, conforming, and narrow in their interests. Those getting lower grades than expected were more independent, self-directed, and culturally sophisticated.231 High-grade and low-test students, as opposed to low-grade and high-test students, were found to be less autonomous and more dependent, with higher social consciences.232 They were also found to have an achiever personality (academic conformity or attitudes associated with grade-making), to score higher on the Survey of Study Habits and Attitudes, and to be lower on the dimensions of intellectual quality and on critical thinking. The high-test and low-grade

228J. A. Conway, "What Are We Rewarding?" Phi Delta Kappan, LI (Oct., 1969), pp. 87-89.


230Pemberton, "The Grade Point Average."

231Ibid., p. 13.

students showed more evidence of broad intellectual interests and a creative temperament.233

Several studies have indicated that there are complex relations between personality and grades. When Spielberger studied anxiety in this context, he discovered that the level of anxiety had no effect on high or low ability students, but that high anxiety was associated with lower GPA for middle ability students.234

A work by Johnson also illustrates some of these complex relationships, for he noted that variables which predict performance in one subject may not predict performance in other areas, and suggested the need to consider nonintellectual factors relative to achievement and ability in the context of other factors, such as age, socio-economic class, parental identification, and other variables.235

In general, factors or personality traits associated with higher GPA include perseverance, control, sociability, conformity, rigidity, insecurity, compulsiveness, and dependence. Traits associated with lower GPA include independence, adaptability, self-direction, cultural sophistication, intellectual quality, and critical thinking. Frequently the relationship between traits and grades is indirect. The interaction between trait and type of school, type of course, interaction of student and instructor (dogmatism), and ability interacting with trait (anxiety), as well as other mediating factors all effect the impact of personality traits upon academic achievement as measured by grades.

233Ibid., pp. 29-30.


Sex

While the literature of grading infrequently reports on the effect of sex on grades, it is an accepted fact that women as a group perform significantly better than men at the elementary and secondary level, and there is no reason to postulate any different relationship at the college level. A study conducted at the University of Delaware does support the superior performance of women, and in every college of the University which enrolled men and women, the average grades for women are from approximately three to six tenths of a grade point higher, both in terms of cumulative grade point averages and in terms of grades earned during the junior and senior years.236

Growth or Improvement

Almost no studies have attempted to consider growth in student achievement during a course, as distinct from absolute achievement, as a factor affecting the assigned grade, though several writers did indicate that this factor was used to assign grades. One person suggested that improvement ought to be used in accordance with the nature of the course, and be weighted about ten percent on a trial basis, until more experience could be gained in the use of this factor, and his formula was presented.237

Grading Outcomes

Since the objectives of grading are multifold and perhaps conflicting, it is difficult to summarize the effects of grading, or even to determine the appropriate categories for conceptualization. It is easily seen that the effects of grades on the academic world are pervasive and complex, just as the determinants of grades are also complex and variable. The analysis of grading outcomes which follows...

236Pemberton, "The Grade Point Average," p. 18.

here will be made according to the summary of grading purposes outlined above; other effects of grading and the effects of the P/F option will also be studied in this section.

Grading Objectives

Extrinsic Objectives

Selection for Graduate School

Undergraduate grades are of critical importance in graduate and professional school admissions. In 1969, when 102 admissions officers were asked to cite selection criteria used at their institutions, they submitted the following statistics: overall GPA, used by 92%; letters of recommendation, by 60%; Graduate Record Examination, by 40%; all other criteria were cited by less than one-third of the respondents. Three surveys of graduate, professional, and medical schools showed that GPA was considered the single most important predictor of success in graduate school. As a consequence of the importance of GPA for this selection function, the relationship between undergraduate GPA and achievement in graduate school has been a fairly well researched topic, just as high school achievement has been extensively studied as a predictor of college GPA.

Most studies exploring the correlation between undergraduate GPA to graduate GPA and/or ratings by faculty have found correlations ranging around the .30's, though one


study indicated no relationship. These usual correlations, though statistically significant, explain only about 12% of the causal factors. In terms of GPA, it means that students with an undergraduate GPA of 3.5-4.0 have a predicted graduate GPA of 3.58; 3.0-3.49 have a graduate average of 3.51; 2.7-2.99 to 3.48; and 2.7 and under, 3.36.

Lannholm, reviewing studies by himself and others, concluded that the best predictor of success versus failure in graduate school was the advanced area test of the GRE; combined with the GRE aptitude test, he said, it was a better predictor than the undergraduate GPA.

In general, undergraduate GPA has some use as a predictor of success in graduate school. A survey of a large number of studies showed that small positive correlations on the order of .30 to .40 are usually found, though the correlation may be higher or lower, depending upon the school.


and the field of study. In some fields of study undergraduate GPA apparently predicts early success better than later success. These correlations explain only approximately 10% to 16% of the variance due to whatever is included in the global undergraduate GPA. That these correlations may be somewhat low in the light of the constricted range is probably due to the fact that most lower ability students are excluded from graduate school and therefore also from these studies.

Selection for Employers

Employers also look closely at grades in selecting new employees. A survey by Dickenson, for example, showed that various factors such as sociability, conscientiousness, drive, training, and judgment were considered important, but that intelligence was the most important in the composite rating for all fields. If it is assumed that employers judge intelligence by grades, then grades are to that extent of primary importance.

When the undergraduate grade point average is used as a selection tool to assist employers, it is necessary that the GPA predict later adult achievement. Fortunately, a definitive report on this relationship has already been written by Hoyt. The report, "The Relationship Between College Grades and Adult Achievement: A Review of the Literature," was published as an ACT Research Report in 1965, and two later articles were published by Hoyt on the same material.

---


A number of problems faced in this type of research were noted by Hoyt. First, research has focused on vocational achievement and ignored most other criteria which could reflect other types of academic success, some of which could presumably interfere with vocational success. Second, the range of academic achievement is constricted, since only college graduates are included, and thus correlations would be somewhat lower. Third, specification of success criteria has frequently been salary, but it is better conceived as multi-dimensional. Differences between groups, firms, and colleges produce various complications in the analysis of relationships. Some professions are higher paying; some colleges have very different standards from others. Fourth, the time frame can affect results. Quick follow-up might be inaccurate as graduates have not had time to establish a record of success. Slow follow-up allows more factors unrelated to college achievement to affect adult achievement. Last, a final problem is the self-fulfilling prophecy nature of the relations of GPA to adult achievement. Even if college or academic achievement has no relationship to adult achievement, as long as employers think it does, and as a result offer better jobs and salaries to those with higher GPA, an artificial correlation between the two will thereby be created. This latter problem, though serious, is formally considered in the literature by Hoyt alone.

Hoyt's actual findings follow. On the basis of seven studies of achievement in business, little correlation to undergraduate achievement was found. Only two of the studies found significant relationships, which were with salary, and the contingency coefficients were approximately .35. Of twelve studies relating teaching effectiveness to GPA, only a few indicated a positive correlation, and in these instances the correlations were low, generally between .20 and .30. Correlation varied, depending upon how GPA was computed (student teaching grade, education grades, all grades) and according to what aspects of teaching performance were being considered. One study not included by Hoyt reported that teachers with GPA's from 2.20 to 2.49 were judged equally successful in the first year or two of teaching as those with

Hoyt, "The Criterion Problem."

Hoyt, "College Grades and Adult Achievement," p. 20.
GPA's of 2.5 or above. In engineering, Hoyt found that four of five studies showed no relation between grades and salary. The study with affirmative results was experimentally questionable. In medicine, eight studies indicated that medical school grades seem to bear some relation to early success, but not to later success; in this area surveys showed that undergraduate GPA was not related.

Five studies concerning contributions to scientific research, among the best methodologically, indicated rather modest positive relationships to college grades, on the order of .25 to .35, and five studies of miscellaneous occupations also showed very slight, if any, relationship between grades and salary. Two studies of non-vocational achievement, such as colleges frequently cite among their goals, (i.e., civic participation, current events information, social activity, cultural interests) likewise failed to indicate any connection with GPA. Nine studies of eminent achievement using various criteria of adult accomplishment found only a modest relationship between such adult achievement and academic grades. In fact, the only positive relationship found was with the amount of additional higher education; indeed large numbers of eminent men came from the lower academic achievement ranks.

---


251 Ibid., p. 30.

252 Ibid., p. 36.

253 Ibid., p. 38.

254 Ibid., p. 40.

255 Ibid., p. 43.
The following conclusion was drawn by Hoyt:

Despite these limitations, however, we can safely conclude that college grades have no more than a very modest correlation with adult success no matter how defined. Refinements in experimental methodology are extremely unlikely to alter that generalization; at best they may determine some of the conditions under which a low positive, rather than a zero, correlation is obtained.256

A survey of college graduates did indicate that higher grades were related to financial success in most fields, though no relationship existed in business.257 Some eight additional studies were reviewed by Jonathan Warren in 1971, and although some evidence was presented both for and against the relation between grades and adult achievement, he concluded that Hoyt's overall conclusion still stood. Previously, Raimi, noting that the research findings suggest that GPA does not correlate very well with later achievement, proposed two possibilities to explain this result: 1) good grades are not well correlated with good academic performance, and 2) good academic performance is in itself a poor predictor of later success in one's occupation.258 In the first case, he said, educators must make appropriate corrections; in the second, he believed there was no problem, since so many other factors are involved that grades should not be expected to correlate highly with adult achievement. In a later section of this 1967 study, Raimi discussed the limited nature of information provided by grades about academic ability and achievement and considered them to be only one of many relatively independent abilities and achievements. Additional reasons he offered for the inability of grades to predict adult success were such factors as IQ, personality, and job experiences, all of which would likely be better predictors of success than grades, which should not, he felt, be downgraded for failing to be what they are not.

256Ibid., p. 45.

257E. Havemann and Patricia S. West, They Went to College (New York: Harcourt, Brace, 1952).

Fencks and Rieseman explained how modest correlations may be a result of other factors:

Even if there were no causal relationship between mastery of academic subject matter and subsequent achievement, we would expect the brighter and more diligent individuals to do relatively well in school, simply because school constitutes a challenge, and they presumably have the equipment to rise to it. We would also expect these individuals to do relatively well in the world of work, for similar reasons. This would produce a statistical correlation between grades and job success even if there were no causal link. In addition, even if there were no causal relationship between the learning measured by grades and success, we would expect students with good grades to be given certain advantages not open to others, simply because so many employers assume a relationship between grades and occupational potential.259

In general, the relationship between GPA and adult achievement is very modest at best, and future studies are unlikely to produce different results. In addition, the relationship should not be expected to be strong given the multidimensional nature of grades and adult achievement.

Reporting Function for Parents

This particular function of grades, though frequently cited in the literature, has not been considered an important justification of grading. Since grades in practice do determine a student's present scholastic success or lack thereof, they must, of course, accurately and effectively inform parents of their son's or daughter's progress. As long as specified grades are a part of graduation requirements, this function will continue to be served by definition, and even non-traditional grades can achieve this purpose to a large extent, since a "Pass," while perhaps less informative to a parent, still is indicative of satisfactory progress in the context of institutional requirements. One might maintain, however, that the information passed on to parents may or

may not coincide with "real" progress. To the extent that grades reflect a variety of factors which are unexplained and may or may not be related to academic achievement (as discussed above under determinants of grades), this reporting function is not well served.

Institutional Objectives

For the most part, the achievement of institutional objectives has not been questioned in the literature. These objectives are achieved almost by definition, however, as they are part and parcel of the grading system: credit according to the rules is based upon a certain grade; honors are given on the basis of grades, and they are frequently defined simply by grade point average; grades not only determine a student's progress and continuance in school, but presumably are also considered fairly accurate indicators of the student's preparedness for advanced work in an area, since the grading criteria, ability requirements, etc., are presumably the same in the advanced course in the same area. Apparently success or a satisfactory grade at the lower level should predict success at the upper level; in fact, grades, even in unrelated courses, tend to correlate fairly well on a semester to semester basis.

The one institutional objective mentioned in the literature but not usually achieved in practice is the measure of efficiency of instruction. As noted above, there are many determinants of grades and sources of variation which make any reliable determination of efficiency of instruction based on grades an impossibility. Only a multi-sectioned course with common measures of student achievement would allow for any valid estimate of efficiency, and this is not the normal case. Dressel and Nelson stated that the concept of a grade as a qualitative judgment rendered by an infallible judge was patently unacceptable, since there was no simple relationship between the distribution of grades by a teacher, the standards of the teacher, and the quality of the teaching.260

In 1970, some 549 undergraduate institutions responded to a survey concerning GPA and its uses. Ninety-four percent

260Dressel and Nelson, "Testing and Grading Policies."
reported computing the GPA. Uses of the GPA included determining honors and ranks, graduation, suspension, dismissal, probation, and restriction.261

The actual achievement of institutional objectives has not been a matter of great contention. The critics of grades, however, questioned the validity and not the achievement of institutional objectives. Ericksen is typical of such critics.262 If grades, for example, are unreliable or do not truly reflect the achievement of desired educational outcomes, then should grades be used to award credit, determine a student's continuation, be the basis for academic honors? The various "ifs" are considered below.

Intrinsic Teaching-Learning Objectives

Motivation for Students

A number of studies have been reported which at least indirectly bear upon the motivating function of grades. Among the conflicting results reported as effects of midterm grades or "in-course" feedback to students are: some improvement, and more by women than by men; improvement by poor and middle students, and poorer work by good students; no difference for F students, and lower performance by D students; consistently good performance by good students; much better performance by A students, poorer performance by B students, and unchanged performance by C students; a maximum effort at study made by students near failing.263

261 Buchman, "Grading Systems."
262 Ericksen, "Grading Evaluation."
Thayer, "Do Low Grades?"
Birney, "Effects of Grades on Students," pp. 96-98.
These results are obviously complex, contradictory, and confusing. The experimental control and quality of some of the studies is questionable, and some of the results listed might be explained by regression to the mean, the tendency of instructors to grade more leniently at the end of the semester than at midterm, and so on. Though such in-course feedback to students apparently has motivational impact, the degree and relationships to such factors as ability level, sex, discipline are unclear.

The effect of academic probation upon subsequent achievement was reported in four articles by Fischer. On a long-range basis, he found, women on such probation showed a slightly higher GPA, and men a slightly higher average for one or two semesters. There also seemed to Fischer to be various interaction effects between probation, sex, and field of study, but the effects of probation in these studies was not noteworthy and might be partially explained in terms of regression to the mean.264

In one study, interested in the effect of expectation upon achievement, some 1400 freshmen were told that their predicted GPA was 0.40 higher than the actual predicted GPA. Results, however, showed no effects upon actual GPA,265 perhaps because the harder than expected grind for expected grades counteracted any positive effects.


Studies of student opinion generally confirm the motivating property of grades, though one very small sampling of students reported the opposite.

Several writers have pointed out the pervasive effect of grading and grades as reinforcers of behavior in the nature and structure of the educational system. Dreeban analyzed the impact of social organization and socialization on children and the values consequently imported, holding that orientation of children to grades and teaching students to appreciate them are important tasks of the school. Only after grades acquire a "sanctioning" quality, Dreeban asserted, can they be used to guide student behavior, since the judging of so many specific activities on a "good and bad" basis, and the grade-symbol reward is not a usual feature of the child's home environment.

Wolff, a strong critic of grades, scored the fact that grading is the dominant educational reality from first grade to college, stating that the student "lives, breathes, and defines himself in a world of grades." He believed that the best of students become fixated on grades and come to seek the symbol as an end in itself, and concluded that if the lower levels of education should give up this particular socialization process—as some are doing—the change would affect the grading scheme in higher education.

Most commentators have accepted the motivating influence of grades. Dressel and Nelson considered exams, as

---


269 Wolff, "Ideal of the University," p. 58.
the basis for assigning grades, to be potent elements in defining the objectives of a course since students studied for course objectives as they are given expression in the examination. Buxton also was convinced that students were strongly oriented to achieve on examinations and concluded: "It is clear that for every stated educational objective of a course there must be a corresponding effort to measure that achievement, and that the students must be aware that such a measurement effort will be made." 

One conclusion of the research on pass-fail grading reviewed above relates to motivational effects of grades. Surveys clearly showed that student achievement (as measured by grades) was significantly lower under the pass-fail option than under traditional grading. The following explicit conclusion by Karlins accurately reflects the results of pass-fail research:

(3) students generally believe they learn more, work closer to their capacity, and are more motivated to learn in a numerically graded course than one marked pass-fail; (4) students believe they participate more fully in a course when faced with competitive (1-7) rather than pass-fail grading; (5) students believe that working for competitively assigned 1-7 grades can be tension inducing, but such grading can also stimulate the student to work harder in his courses.

There are, of course, different types of students, and grading may not affect all in the same manner. Korn distinguished three types, basing his judgment on what the curriculum meant to each group: (1) those interested in getting good grades; (2) those concerned about a career; (3) those pursuing intellectual interests. Showing that these groupings had different personality characteristics, he concluded that there are profound differences in the ways


271Buxton, College Teaching.

272Karlins, Kaplan, and Stuart, "Academic Attitudes and Performance," p. 44.
students learn and in their underlying patterns of motivation for learning.273

Of critics of the motivation function of grades, a first group maintained that grades are not good motivators; a second admitted, at least implicitly, that grades were strong motivators, but believed that the behavior thereby motivated was not consistent with educational objectives. Trow, a member of the second group, stated: "Unfortunately, the motivation induced by letter grades is not the highest and noblest, which may be one reason why professors, as well as students, are often willing to dispense with it."274 Other members of the second group, such as Erickson, equated motivation provided by grading with coercion,275 and Pemberton, who severely criticized the motivation of grading in providing only extrinsic rewards and punishments which will not basically affect a student's will to learn. Pemberton concluded:

The argument comes down to this: grades and threats of flunking a student are the only whip that teachers have left for keeping lazy and unruly students in line. This is a losing battle. If one cannot show students that he is an authority on something, he will not long be able to exercise authority over them.276

The relationship of grades to teaching and learning emerges as the primary issue in the debate on grading. Indeed, perhaps the most critical issue for teaching and learning is the problem of student motivation and grades. The ratio of sound to sound research is greater here than for any other grading topic. Very little substantive research has been undertaken and Birney noted that what has been done has related for the most part to studying grades only as measures of achievement, ignoring them as stimuli to behavior.277

273Korn, "Differences in Student Responses."
274Trow, "Grades and Objectives," p. 87.
Pass-fail literature, for example, clearly shows the motivational effects of grades. Literature on grading has not, however, demonstrated the feasibility of other alternatives. It has shown the complex nature of motivational effects, but the exact relationships are unclear, and empirical outcomes are too often contradictory.

In summary, grades do have important motivational effects upon students, and through the focus provided by examinations grades guide their activity and efforts. The motivational effects which are complex, depend upon sex, subject matter, and probably upon other personality traits; defenders of traditional grades go so far as to argue that student effort would be reduced without grades, yet adverse critics maintain that the activities stimulated by grades are not educationally sound. Some of these motivational aspects are considered below.

Feedback to Student on Progress

Little or no research has been directed toward determination of traditional grading's effectiveness in achieving this objective. Several commentators have suggested that a grade provides only very limited information. Marshall believed that grades basically indicated "yes" or "no," and did not provide substantive information about what a student was doing well or poorly and why. They told a student only where he stood in comparison with others, he said, and not what was right or wrong, and he added that since grades were easily accepted because of their vagueness, they could flexibly mean whatever one wants. "A" meant excellence, but excellence was not defined, so everyone was happy.278 Axelrod believed grades are an inadequate measure of achievement and had no consistent meaning.279 His study of several thousand students indicated that students did not agree that grades provided feedback indicating what material has been learned.280 The research on midterm-grades cited in the previous section also was inconclusive.

278Marshall, Teaching Without Grades.
279Greene and Hicks, "Normal Distribution?" pp. 135-136.
Beyond the issue of what grades may or may not express in addition to achievement, there is a serious problem of definition in assessing this objective. It is clear that the grade or symbol given to a student and his work at the end of the term is a final, comprehensive, and multidimensional judgment which tells the student in general how he did. By and of itself, however, it provides little feedback to the student to facilitate the learning process or to assist the student in understanding areas of strength and weakness, yet if the definition of grading is broadened to include the processes of testing, evaluation of class participation, and judgment of term papers and other projects, all of which are used to determine the final grade, then a strong case can be made on behalf of grades. Grading critics respond that all these processes are possible without ever assigning a final grade, and, in effect, are assuming the narrow definition of grading, that is, the assignment of a final grade. On this basis it is maintained that the assignment of a final grade and the resultant implications impair all the other feedback processes which assist learning. Defenders use the broad definition and strongly argue that if it were not necessary to assign a final grade which could be defended to the student, if need be, then instructors would very soon become careless and neglect these feedback procedures, with harmful effects upon learning.

In addition, the manner of testing minimizes the possibility of useful feedback on progress to students. Since final exams and many other course examinations are summative or final on a specified content area, the student has no opportunity to reflect upon the grade and his lapses in knowledge in order to show mastery at another later date. If students were permitted to repair errors and study further, the success of the grading system, broadly defined, in providing feedback to students, might be much better.

Both positions appear to be partly true. Under pass-fail grading, for example, there is some evidence to indicate that evaluation by instructors is less careful, since it will not affect the grade. Critics have also expressed clear concern that the spectre of a final grade motivates the student to attempt to respond to the bias of the instructor and to avoid any demonstration of ignorance, neither of which reactions is beneficial to learning.
Feedback to Instructor on Students' Progress

The same concerns discussed above regarding feedback to students apply here. To the extent that grades signify things other than achievement of intended academic goals, they cannot provide adequate feedback to the instructor. Likewise, any grade defined narrowly as the final grade fails to fulfill this objective. The various evaluation processes culminating in the final grade do achieve this function. The same conclusions apply here as in the case of feedback for students.

Other Outcomes

Preparation of the Student for Life in a Competitive System

Some defenders of traditional grading have argued that students will engage in a competitive world after graduation and that, over and above any other considerations, grades are valuable in preparing students for this world. They do not present this as a major justification for grading.

This reasoning rests on some dubious assumptions. First, the competitive world, at least as students could relate it to grading, may not exist: few persons work in situations where a series of uniform tasks is required and where they are ranked consequently on a relative or normative basis. In fact, in a number of fields, such as teaching at the elementary and secondary level and most union work, non-competition is the basic fact of life. Undoubtedly competition is a factor in many areas, but the competition between attorneys, doctors, executives is not very similar to what results from grading. Even if some similarity is present, the assumption that the competition of grading is the best preparation for it is also unsupported. A study of several thousand students showed that opinion was fairly evenly divided between those who thought grades offered a realistic model for the competitive nature of life and preparation for life after school and those who did not.281

281Ibid.
In fact, the opposite argument could be made: college is the time to explore, to make mistakes, and to learn from them. Yet Bloom points out that nowhere else is an individual judged so frequently or in such detail as in education for the student judges himself by standards of self, teacher, peers, and family. Since success or the lack of it determines positive or negative self-concepts, in a competitive or ranking grading system some students by definition must be unsuccessful. Those in the lower brackets will, therefore, approach additional learning tasks with more reluctance, less interest, and less confidence. Research indicates that students at more selective institutions, on receiving lower grades than students of similar ability at other institutions, on the contrary, are apt to be discouraged from graduate study.

If renewed interest in learning would follow, mistakes for which a student was penalized could well lead to a higher grade. Non-competition in an educational environment, which does not finalize study with the final grade in a course, may well be a better form of preparation for competition as it will be experienced later. One writer suggested that while grading reinforced the archaic notion of competition, it is cooperation which in reality is an important part of life. In any case, argumentation on both sides is speculative and unsupported by research so it is perhaps prudent to avoid this argument in defense of traditional grades.


Charles E. Werts and Donovan J. Watley, "A Student's Dilemma: Big Fish Little Pond or Little Fish Big Pond" (National Merit Scholarship Corp., Evanston, Ill.: 1969). ED 029 559.

The well known Hazen report, *The Student in Higher Education*, saw competition as an obstacle in the area of student development since testing and examinations require the student constantly to prove himself, thus eliciting feelings of inadequacy.285 The report recommended in consequence that competition should be reduced.286

Sociological Results of Grading

There are two important publications which explore student culture and the social attitudes and nature of the effects of traditional grading. Becker, Geer, and Hughes in *Making the Grade* reported on the collective modes of action and perspectives of students toward grades at the University of Kansas.287 Wallace in *Student Culture* reported his in-depth sociological study of a private liberal arts college concerned with student socialization, grade orientation, and GPA.288

The first study showed that students perceived three areas of college life: academic, extracurricular, and personal relationships. In the academic area, the college, on a more or less unilateral basis, established the conditions for student activity, which were the same for all, thus setting the stage for a similar and shared perspective. Attitudes and interpretations will grow out of shared experiences, which would include definition of the situation, criteria of judgment, and appropriate activities, a result called "the grade point perspective" by the authors.

In general, success in the academic world has been geared to grades. All other institutional awards assumed successful grades as a precondition, and to the student a


286 Ibid., p. 61.

287 Becker, Geer and Hughes, *Making the Grade*.

"mature" student who is "doing well" is one who can successfully meet that requirement. Therefore, though other intellectual interests and activities may be pursued, should they conflict with the activities necessary to pursue adequate grades, they must be sacrificed. Students expected to get "good" grades and to do whatever is necessary to obtain them, not expending energy on other goals until the minimum "good" grades were secured. The success of this pursuit of satisfactory grades is perceived by students as a sign of one's maturity and self value. Becker concluded:

Students generally accept this perspective whether they like it or not because it is realistic, and is accepted by the peers. There may be private reservations, but this is the acceptable practice.289

Grades can be compared to a campus currency, according to Becker, being the chief institutionalized valuable and the only one formalized in an officially-administered system applying to all students. In the context, grades serve in a sense of money, and neither scholarship, nor intellectual growth, while considered valuable, is held in high common regard as are grades.290

When the learning process conflicted with grades, grades were dominant. The student with this estimate of grades in mind therefore approached each class looking for clues which will enable him to get the desired grade, that is, to perform as the instructor wants him to and was prepared to exchange certain performances and activities for the desired grade. The instructor explicitly or implicitly made a contract with the student, whenever he indicated course requirements--on a course syllabus, perhaps, by the type of behavior rewarded in his class, most especially on course tests. The students looked for subtle inferences in the instructor's behavior to determine what kind of achievement is especially rewarded. Believing that instructors vary greatly in their demands for academic achievement, they felt that they must study each one's emphases. Does he stress the importance of punctuation and spelling and the expression of one's own viewpoint? Does he desire creativity?

289 Becker, Geer and Hughes, Making the Grade, p. 35.

290 Ibid., p. 55.
Do his tests require specific information, logic, or what? In general, "This view of classroom interaction defines the teaching situation as an exchange of rewards for performance, rather than as some kind of 'educational process'."  

Faculty were judged by students to a considerable extent on how difficult or easy they make it to get a "good" grade. Students looked for two sorts of information in planning their grade-getting activities, the formal requirements stated by the instructor and the informal requirements hidden perhaps in subconscious cues given by the instructor and picked up from other students. In preparation, students adopted various strategies depending upon the type of demands made, i.e., essay exam, objective exam, etc. Because grades were so important, students wanted them to be awarded in a fair and serious way, equitable to all students, objecting to arbitrariness and random variation. They also objected to the use of different procedures for marking in sections of the same course as well as to a system in which a high B that is almost an A counts the same as a low B that is almost a C. They were upset when the system was changed, so that unexpected material was included in a test, or when a different method of testing or determining grades was used, so that the students' "game plan" was outdated or incorrect. 

Students undertook a variety of activities to get good grades, which activities included doing the assigned work, cheating, relying on files, having other students assist in work, ingratiating oneself with the instructor. Projecting their semester grades, they obtain a basis for their decisions about the allocation of the time and effort which will be needed to get the maximum results.
Although students may originally be motivated to work when required to do so to get good grades, such application to learning may result in the habit of only working under such conditions. Faculty generally believed that grades resulted from the combination of a student's ability and effort, yet many of the student's grade-oriented activities were completely outside this supposition. Becker stated:

The conventional academic system has a built-in irremediable conflict. Professors do not like students to seek information of the kind we have described; that seems to them a misdirection of effort and curiosity that might be better devoted to the content of the course. But students are constrained to engage in information-seeking activities by the importance of grades and the ambiguity of professorial statements about how they may be achieved.

In the same spirit Becker continued:

The import of this, from a conventional faculty point of view, is that, given the importance of grades and the total control by faculty over the terms of their distribution, students cannot act as autonomous intellectuals, cannot pursue learning for its own sake, but must seek information on faculty behavior, present and prospective, before they can plan what they will do. In this sense, the relationship of subjection works against the commonly stated faculty goal of training students to be intellectually free and self-directing, and does so as a matter of structural logic rather than as a matter of human frailty, or student or faculty incompetence.

The authors concluded that the grade-point average perspective was frequent, widespread, and collective among students, although understandably such a view of their academic

296 Ibid., p. 97.
297 Ibid., p. 98.
298 Ibid., pp. 86-87.
299 Ibid., p. 91.
work was unlikely to please their faculty members. Nevertheless, this attitude was a reasonable and inevitable result of the way colleges were presently functioning.

While Becker's research was based upon the method of participant observation, Wallace relied heavily on measurement of student orientation as compared with grade achievement at various time intervals, and on comparison of grade orientation to grade achievement. A detailed study of attitudes during the beginning of the year, according to Wallace, showed a great amount of socialization of freshmen in the first seven weeks of college and a continuing change afterward in terms of decreasing grade orientation. Grade orientation (GO) of freshmen, Wallace's study shows, tended to decrease and stabilize at the same level as that of upper-classmen, and to fall below faculty expectation. The greater the contact of a freshman with non-freshmen, the more congruence was found with the upperclassman attitude toward GO. There were indications that parents had little influence on GO, that peers exercised negative influence, and that postcollege plans positively influenced it. A need for social acceptance led to lower GO. Entering freshmen first thought high grade orientation was valued. When they quickly saw that it was not, they lowered their value of GO to gain acceptance.

The scheme below outlines the relationship found by Wallace between pre-college ability, GO measured in November, first semester GPA, GO measured in April, and second semester GPA. An arrow is interpreted as "leads to." Since the relationship between ability, grade orientation, and actual achievement is complex, all possible combinations are not set out below. It is clear, however, that a student's attitude toward grades, acquired from and in the context of peer relationships, had significant effects on his activities and actual grade achievement.

300Wallace, Student Culture, pp. 50-51.
There is a seeming contradiction between the studies of Becker and Wallace, in that the former implied that the student will do anything for grades and the latter suggested a reduced grade orientation after coming to college. The resolution lies in the use of this term "good" grade by Becker. Good is defined on an individual basis by each student and may be C in one instance and A in another. Grade orientation refers to absolute level of aspiration. While students lower their grade orientation upon coming to college, they will undertake whatever activities necessary to achieve their satisfactory grades.

Other writers have also noted the centrality of grades in a social context, without studying it so thoroughly. Dressel and Nelson claimed that colleges, through the uses of grades, cultivated students' preoccupation with grades even though professors are impatient with that attitude.301 Katz reported that many students saw grades as the central reality of academic life, since they determined future rewards and supply a measure for the student's self-esteem.302 McKeachie

---


indicated that grades were a most potent motivational device: they held the key to graduate school and jobs; remaining in college depended on them; they served the student as an "expert's" appraisal of his success, if he was interested in learning. He stated:

Because grades are important to them, students will learn whatever is necessary to get the grades they desire. If we base our grades on memorization of details, students will memorize the text. If they believe our grades are based upon their ability to integrate and apply principles, they will attempt to do this.303

Just as a grading sociology has emerged among students, a web of values and attitudes was prevalent among faculty. Some faculty in order to maintain academic standards were hard graders. Others were student-centered, and administered good grades liberally to encourage students' dedication to learning. Different disciplines, and groups of related disciplines, often inculcated different values. Science and math faculty, for example, tended to be rigorous while faculty in the social sciences and fine arts generally were not. The instructors' philosophy of grading in freshman courses differed from that used in grading upperclassmen. Katz commented that the social structure and attitudes generated by faculty interaction have not been well studied.

Student Opinion

Student opinion has been a part of the discussion of many grading topics considered above and attempts have been made to justify various positions on this basis. Unfortunately, however, the quality and validity of most surveys of student opinion are uncertain at best since each survey instrument is generally the creation of one particular writer, little effort is made to check reliability, and items are such that the response may be biased. The statement, "Despite instructor's insistence that they do not teach 'facts,' most grades are based on tests which are primarily factual

in content," is only one of the more flagrant examples of such bias.304

In any case, a number of the larger and more reliable studies are presented now and provide a good review of the findings. Students overwhelmingly endorsed the traditional grading system in 1937. About 80% were satisfied. Some 15%, mostly those receiving low grades, believed that grades were too low. Five percent judged them to be too high.305 More recent studies of the 60's and 70's showed much greater dissatisfaction,306 indicating that 50 to 70% of students interviewed expressed serious concern about the present grading system. A third to a half of those responding considered grades a major worry and admitted that because most grades are based on tests that are largely factual, grade emphasis was restricting their study to test material. Grading emphasis was also seen as promoting cheating and motivating students to conform to the instructor's views, on tests and in the classroom. Grades were viewed as the


Katz, "Four Years of Growth."

central reality of academic life in determining both rewards and self-esteem. About half the students believed that grades stimulated better work, but only work likely to be closely related to specific course objectives. Merely a minority of students thought that grades provided useful feedback.

Some studies reported that better students were more favorable to grading and some that they were less so. One writer concluded that students wanted high grades but did not want to compete for them and rejected the present method by which achievement is measured and rewarded.307

One question on grading included each year in a national survey of entering freshmen sponsored by the American Council on Education asked whether grades ought to be abolished in college. In 1970, 42.9% were reported in favor of dropping grades,308 while in 1973, 38.5% were in favor.309 The most recent survey indicated that the number wishing to abolish grades has dropped to 28.2%. While it may encourage traditionalists that a minority and not a majority answered in the affirmative, and that a slight decline has occurred in that number, it must still be disconcerting to many educators that about two-fifths of all freshmen just entering upon their collegiate work were in favor of taking the drastic step of abolishing grades in college, results far from the rather contented results reported in 1937. A number of studies with smaller samples generally reported the same findings.310

---


In general, studies of student opinion report serious discontent with traditional grading systems and favor some modifications. The majority of students, however, is not in favor of doing away with all grades, since grades are generally considered to be motivating, but most agree that various negative aspects, such as the pressure on students to conform, the encouragement of cheating, the basing of grades on extraneous factors (which were noted above), should be considered. Most studies suggest, too, that the quoting of percentages of students with particular attitudes should be viewed with some doubt, since the reliability of such reports and the bias effecting the formulation of items used as proofs is undetermined.

Faculty Opinion

In studies and surveys of faculty opinion, which have been smaller and less frequent than those relating to students, the same concerns about validity, reliability, and possible bias are applicable.

Faculty opinion reported in the Berkeley study, noted just above, was based primarily on comments and letters received by the authors. In general, some problems with traditional grading were noted, but any other system was seen only to exacerbate these problems, leading to the conclusion that if formal grades are subjective, any other system would be more so. Some few indicated that the measurement of a student against his peers was a salutary thing.311

Dysfunctions of Grades

There is no simple way to organize all literature on grading into distinctly different categories. One may begin with the large volume of such literature which, for a variety of reasons, argues against grading, noting that while much of it simply represents personal opinion, some does present carefully reasoned argumentation. In any case, literature dealing with the dysfunctions of grading is an important segment of the total grading literature since it frequently presents plausible points for consideration and suggests

---

311 Education at Berkeley, pp. 95-96.
areas for needed empirical research. It is, therefore, reviewed here by categories of concern.

Distortion of Ends and Means

A frequent complaint against traditional grading is that the search after grade symbols has, in effect, replaced the end of undergraduate education, which is knowledge. Pemberton thought this to be an ethical matter, declaring, "There is ethical confusion when grades are perceived by students as ends rather than means."\(^{312}\) Thorndike, in his review of grading for the 1969 edition of Encyclopedia of Educational Research, stated, "Marks are little, if at all, related to achievement of the central and important objectives of the educational program, and focus attention upon false and inappropriate objectives,"\(^{313}\) adding that grades give a distorted education value pattern, tending to make appearances, rather than substance of learning, the important thing. Others expressed the same problem in their own words,\(^{314}\) and mentioned that broader educational goals, such as maturation, liberation, ability to analyze, were not included in grade reports.

According to Raimi "... the most telling criticism of all grading systems is this: that the incentive and discipline they foster are incentives to beat the grading system itself (rather than towards scholarship), and discipline in the direction of safe conformity (rather than in the habits of learning)."\(^{315}\) Marshall made a similar complaint, holding that grades and scholarship were not synonymous, and

\(^{312}\)Pemberton, "The Grade Point Average," p. 2.


\(^{315}\)Raimi, "Examinations and Grades," p. 311.
that grades encouraged the tendency of students to please teachers. Other commentators316 and also students317 have repeated the same idea.

A broader aspect of this concern has been developed by Holland and Richards who found no relationship between achievement as measured by high school grades, and between ability as measured by ACT scores and by seven areas of non-academic achievement: namely, leadership, music, drama and speech, art, writing, and science activities. According to these critics, academic ability must be considered as only one of several relatively independent dimensions of talent. They conclude:

Our present findings, however, suggest that the emphasis in colleges and universities on academic potential, a relatively independent dimension of talent, has led to neglect of other equally important talents. If academic talent had a substantial relationship with vocational and other non-classroom achievement, then this intense, pervasive concern with academic potential would be less disturbing. Unfortunately, college grades are generally poor predictors of real-life success and are at best only inefficient predictors. Since a college education should be largely a preparation for life, both in the community and in a vocation, we need to examine grading practices. Currently a college education is mainly preparation for more education in graduate school.318

The basis of all such criticism is that desired learning and grades are not congruent. It was noted above that most students (for better or worse) are strongly motivated to achieve satisfactory grades. To the extent that grades and learning do not coincide, and the literature cited


Simon, "Grades Must Go," pp. 8-10.


above and below give some weight to this view, such criticism is justified. The question is then reduced to whether congruence can be increased to satisfactory levels, or whether grading should be scrapped. Answering that is not easy, for there is insufficient evidence to demonstrate the precise degree of incongruence, but there is ample opinion reported in the literature to warrant concern.

**Limited Scope of Information Provided by Grades**

The scope of information conveyed by grades and the related collection of data used to assign grades are closely related to the previous criticism of grades. To the extent that relevant data pertaining to desired educational objectives is not used and to the extent that grades do not validly represent these objectives, grades are dysfunctional.

Several writers specifically suggested that grades are an inadequate measure of educational growth.\(^{319}\) Axelrod believed this and reported that the creative student did not respond well to the traditional grading system.\(^{320}\) (The relationship of grades to creativity was discussed above). In the context of the very low correlations between grades and adult achievement (also discussed above) Hoyt noted that GPA was the only assessment made of educational progress, and that the variety of purposes included in the typical college catalog, such as self-development and initiative, were not

\(^{319}\) D. R. Brown, "Non-intellective Factors and Faculty Nominations of Ideal Students," Report to the College Entrance Examination Board of a Pilot Study (Mellon Foundation, Vassar College, 1959).

\(^{320}\) Axelrod, "The Creative Student."
assessed.\textsuperscript{321} The same argument was repeated by Dressel:

A variety of objectives have been posed for general education. Frequently the achievement of agreement on objectives becomes an end in itself. Little effort is made to actively relate courses to objectives. Few instructors worry about implementing many or any of the stated objectives as such.\textsuperscript{322}

Much of the problem arises from the need for information to assign grades, and the type of information and manner in which it is collected. Axelrod stated that the need to collect "hard" evidence of achievement determined to a significant extent the kind of assignments, tests, etc., which were administered, i.e., materials which can be judged objectively, to rank students.\textsuperscript{323} Dressel concluded that where stated objectives, such as integration, interpretation, etc., were not the implicit objectives (those actually measured on any form of evaluation) or differed from them, the student will work toward the latter. For example, whatever the theoretical goals of a certain course may be, if a course emphasized memory in the evaluation process, the student will concentrate on memorizing.\textsuperscript{324}

From the initiation of objective tests, critics have noted their peculiar contribution to this problem.\textsuperscript{325}

\textsuperscript{321}Donald P. Hoyt, "College Grades," pp. 70-75.


\textsuperscript{323}Axelrod, "The Creative Student."


\textsuperscript{325}Edward S. Jones, Comprehensive Examinations in American Colleges (New York: McMillan, 1933).
Specific concerns included the emphasis on memorization of factual detail and the neglect of higher order levels of thought and problem-solving, (whether intended or not), the assumption of only one correct answer, and the lack of concern for long-term memory and integration.

This literature points out an area of higher education toward which little substantive research has been directed. For example, a random sampling of test instruments or other methods of evaluation used in various courses might be made. Items of the test could be classified according to a system such as Bloom's taxonomy to ascertain the extent to which memorization of factual matter is emphasized, and then the method of evaluation and the content of the evaluation process should be judged for congruence with course objectives.

In any case, these critics of grading made a very valuable point in explaining the specific ways in which grades may be allowed to usurp the ends of education. Dykstra, who agreed with the above critique, believed, in addition, that college instructors disliked grading for a variety of reasons and were doing an increasingly poor job of evaluation. Since it is difficult to measure most goals, he said, tests were usually reduced to the measurement of content material. He stated:

Given the importance that is attached to professorial appraisals it would seem highly desirable for those in higher education to direct their energies toward improving evaluation, rather than seeking to avoid this admittedly difficult and sometimes unpleasant responsibility.326

---


Chaotic Grading Criteria

The determinants of grades were discussed above, where it was indicated that a variety of factors, such as diverse abilities, creativity, student-faculty interaction, personality traits and sex have a significant impact on assigned grades. It is not necessary here to review this research except to note that a significant number of commentators on grading have considered this use of a variety of criteria as a very serious deficiency. These persons include Ebel,327 Crooks,328 Elbow,329 Trow,330 Erickson,331 as well as many others, and this general concern was aptly summarized by Axelrod:

It may signify solid achievement, or it may not signify achievement as much as promise. It may represent incipient genius or muddleheaded independence. It may be a penalty to the brilliant student for work inattentively done or a reward to a mediocre student for work conscientiously done.332

There is no research to demonstrate to what extent the typical college grade represents pure achievement as opposed to a variety of other factors; indeed, such research may not even be possible, since achievement itself is not carefully defined. Therefore, criticism of grading deals not only with current practice but implicitly raises the issue of whether systematic criteria for grading can ever be employed.

327Ebel, Measuring Educational Achievement, p. 401.
Negative Motivation

It was concluded above that though grades do have important motivating properties for most students, at the same time results are exceedingly complex. A number of critics, though accepting the idea of motivation, nevertheless saw this aspect of grading as essentially dysfunctional. Trow declared: "Unfortunately, the motivation induced by letter grades is not the highest and noblest, which may be one reason why professors, as well as students, are often willing to dispense with it." Trow, "Grades and Objectives," p. 87.

Hodgkinson admitted that grades are a motivating force, but he viewed this motivation as negative and extrinsic, being based on fear. Hodgkinson, "Pass-Fail and the Protestant Ethic," Chronicle of Higher Education, VII (Dec. 11, 1972), p. 8.

Goodman believed that grades promoted cramming and passing, rather than learning, and were not a desirable form of motivation. Paul Goodman, "In What Ways Does the Present Marking and Credit System Inhibit or Promote Learning?" Current Issues in Higher Education, No. 5687 (1964), pp. 123-125.

Gold and others argued that students need to develop intrinsic motivation to learn, but that since grades were extrinsic, learning tended to stop after graduation. Gold, Reilly, Silberman, and Lehr, "Academic Achievement," pp. 17-21.

Marshall and Pemberton maintained that students may even achieve more without grades, given the freedom to learn and to enjoy learning.

It is clear from the review of pass-fail grading that one effect of minimizing grades is a greater freedom for the student to do his or her own thing, independently of the judgment of the instructor, but critics have not been able empirically to show that actual student achievement will be greater without grades. Whether a student is ultimately...

333 Trow, "Grades and Objectives," p. 87.
337 Marshall, Teaching Without Grades.
harmed in some way by pursuing extrinsic rewards, or whether a student can pursue intrinsic as well as extrinsic grades, needs to be further explored.

Faculty Role Conflict

The role conflict that exists between the teaching function and the responsibility for judging and assigning grades to students has been pointed out on several occasions in the literature. Raimi stated that the roles of helping the student and of judging the student were conflicting. Marshall, a severe critic of traditional grading, believed that instructors tended to be concerned about rating students rather than about helping them, contending that the instructor feels a greater need to be able to defend a grade than to know what a student is learning and why. On a survey of student opinion, Burke noted student comments to the effect that traditional grading made the instructor a critic instead of a helper, students admitting that they are unwilling to be wrong, to disagree with the instructor, and would rather be safe. Students also saw grading as promoting an adversary-relationship between themselves and the instructor.

This matter is worthy of further study. Is the reluctance of students to expose their ignorance—which is one facet of sound education—caused by grading or by other factors? This research would not be difficult in a multiple-sectioned course where an instructor might be responsible for assigning grades in one of his sections and an outside examiner in another, for in that situation the openness of students and instructor-student interaction could be studied without great difficulty. While it seems likely that grades do have a great impact on this relationship, and that a role-conflict may exist to some degree, the precise nature and extent of such results of grading is undetermined, as is the impact of grades upon the level of student achievement.

340 Marshall, Teaching Without Grades.
342 Simon, "Grades Must Go," pp. 8-10.
Compartmental Mentality

A general but not infrequent criticism of grading is the emphasis placed on bits and pieces of knowledge under a lockstep system. Raimi described grades as being based on bits of knowledge to be learned under a rigid system in a specified period of time, in consequence of which the student cannot leave and return to a subject according to interest. In two articles Douglass argued that units of credits and grades were in themselves alien to education, averring that they were convenient for administration but fragmentized learning. He leaned rather toward comprehensive examinations not tied to particular courses. However, as was seen in reviewing the history of grading, the elective movement doomed higher education to fragmentation in this sense.

Goodman expressed well the thoughts of the opponents to traditional grading in this area:

Lastly, the competitive grading, the credits, the lockstep scheduling, and speed-up are all part of the cash accounting and logistic mentality that is exactly what we do not need in the automated future, when most of the serious work in life will, or should, be concerned with community culture, citizenly initiative, worthwhile leisure, and social service. These are not gradable and cannot be subdivided into credits.

Negative Self-Image by Student

Some critics have argued that the judgment process of grading can result in a negative self-image in various ways. Pemberton reported that gifted students at selective

344 Douglass, Teaching for Self-Education.
345 Goodman, "Marking and Credit System?" p. 124.
institutions receiving average grades lowered their aspiration as compared to similar students at non-selective institutions, and that underachieving students tended to have lesser aspirations than overachieving students, even though the Graduate Record Examination showed the general knowledge of the former is higher. Simon was concerned that grades debased the student's estimate of self-worth, believing that since a student has little to measure himself against, identity to a great extent rested upon the GPA, a result which constrains identity and self-concept to the confines of a transcript.

Arguing that over the years success or its absence determined positive or negative self concept, Bloom concluded that grading, which implied that most students will attain less than mastery according to the grading system (since most students do not receive A) led persons to expect some lack of achievement and consequently to approach additional learning tasks with increasing reluctance, with less interest and confidence. Success, according to Bloom's study, breeds success, and failure breeds failure.

It is clear, from the literature cited in regard to the sociological effects of grading, that low grades do have a great impact on the student's self-perception. Students are expected to "make it" in achieving some minimally acceptable grades. Surveys indicated that students who received negative midterm grades were discouraged rather than motivated to improve.

In fairness it should be conceded that if poor grades can lower self-esteem, then the converse is also true, that receiving good grades can enhance the self-image. A total pass-fail system or a non-grading system would, by definition, eliminate positive effects along with possible negative effects. While change in self-image resulting from poor grades has not been empirically researched, it seems obvious that for some students poor grades may have an undesirable effect.

---

346 Pemberton, "The Grade Point Average," p. 4.

347 Simon, "Grades Must Go," pp. 8-10.

348 Bloom, "School Achievement."
Other Dysfunctions

A multitude of possible undesirable effects of grading have been pointed out in the literature. A partial sampling includes the following: grades allow the instructor to be autocratic and do not require him to explain his evaluation; grades contribute to impersonalization; letter grades which are ill-defined result in careless evaluation; grading creates undue anxiety among students; grades are given too much emphasis as the criteria of success while other measures of success are ignored; grades may present a false impression of objectivity and lack validity. Adding to the list, several writers declared also that grading corrupts the learning process, promotes unhealthy competition, furthers a game in which students try to beat the statistics.

349 Marshall, Teaching Without Grades.
350 Axelrod, "The Creative Student."
351 Dykstra, "We Reluctant Appraisers," pp. 61-62.
354 Troyer, "Grades Have Gone?" pp. 542-555.
357 The Student in Higher Education. Simon, "Grades Must Go," p. 50.
and idiosyncrasies of the instructor, and breeds conformity. One vocal critic, who may be taken as an example of the writers whose broadsides are occasionally leveled at traditional grading, is Battersby, who condemned grades on many counts. They were inaccurate, he said; poor grades were rarely given for fear of harming students; grades were too much determined by non-academic factors such as social and economic pressures; grading encouraged elitism, paranoia, and insincerity among students and teachers; grades missed the reality of learning and tended to "type" students. They are distorted unknowingly, he concluded, by personal bias and by the idiosyncrasies of the system, yet they required perfection from students and rendered mistakes and occasional lapses, which should be learning experiences, into disasters.

A number of these concerns were discussed earlier, directly and indirectly. The precise import of many of these problems is difficult to specify. Feldmesser, for example, in defending grades stressed the fact that there was nothing wrong with extrinsic rewards and noted that the critics of extrinsic rewards generally insisted on receiving extrinsic rewards for their own work; he also suggested that some anxiety and competition were very healthy. But pro and con opinions on these topics will never resolve the problem. Rather, the issue of grading must ultimately be based on its contribution to the educational outcomes desired and on the ability of grades to discharge the responsibilities exacted by the publics which support higher education.

358 Milton, "What It Is... I Measure I Do Not Know."
359 Axelrod, "The Creative Student."
Marshall, Teaching Without Grades.
Simon, "Grades Must Go." pp. 8-10.
360 Battersby, Typical Folly.
Summary

The literature indicates various outcomes of traditional grading. Writers agree that the selection function for graduate and professional school is served moderately well by grades and that they also report fairly satisfactorily on progress in school to students and parents, and to the institution for various internal purposes. Most students, the consensus shows, are strongly motivated by grades to perform as the instructor expects and to work harder, but whether or not this incentive is dysfunctional is disputed. That there are admitted deficiencies in the quality of feedback to teachers and students for purposes of teaching and learning in present practice is generally accepted, yet without grades the situation, critics say, would likely be considerably more serious.

Another finding is recurrent in the literature: grades have significant effects on student values, peer relations, and self-image. A social structure exists with respect to faculty where one's attitude toward grades is closely related to teaching style, relationship to students and peers, and to field of study. Students, most surveys found, are strongly in favor of the pass-fail option and a significant minority, decreasing in recent years, is completely opposed to any grading.

As a result of studies, a variety of dysfunctions resultant on grading have been pointed out: the motivation of students to pursue trivia, to conform, and to overvalue competition, for example, but the validity of these concerns generally has not been well established.

The limited pass-fail option, which has received a great amount of attention, has not fulfilled its stated purposes, and tends to result in lower student achievement.

Much of the literature on grading outcomes is argumentative or opinion-oriented in nature, for though a fair amount of correlational studies have been completed, as have some descriptive studies, sound experimental studies are practically non-existent.
Technical/Theoretical Issues

Most of the literature reviewed to this point has been concerned with examining various actual and proposed grading systems, what factors tend to determine grades, the purposes of grading, and grading outcomes. The theory of grading, evaluation, and technical issues such as reliability, criterion and normative reference, are considered in this section in the context of this information.

Systems and Evaluation

Systems

Grades and grading systems are clearly related to many other elements within higher education and in the larger society, yet although grading is significant, it is only one dimension of the teaching-learning process of the curriculum and of higher education. Higher education is in turn supported and constrained by the larger society. In fact, each unit and subunit interacts with, constrains, and facilitates the larger units of which it is a part and the smaller units of which it is composed. Because of these real interrelationships, grades must be viewed within the larger perspective of teaching-learning, curriculum, higher education, and even society, as well as the "smaller" perspective of instructor and student.

This wider approach has been called the systems concept, which is "primarily a way of thinking."\textsuperscript{362} As a management concept it has been described as follows:

It provides a framework for visualizing internal and external environmental factors as an integrated whole. It allows recognition of the function of subsystems, as well as the complex supersystems within which businessmen must operate. The systems concept fosters a way of thinking which, on the one hand, helps the manager to recognize the nature of complex problems and thereby to operate within the perceived environment. It is

important to recognize the integrated nature of specific systems, including the fact that each system has both inputs and outputs and can be viewed to recognize that business systems are a part of larger systems—possibly industry-wide, or including several maybe many companies and/or industries, or even society as a whole. Further, business systems are in a constant state of change—they are created, operated, revised, and often eliminated.\textsuperscript{363}

Higher education also functions in relation to many other systems and must be understood in the context of interrelated systems. Its outputs of knowledge and skilled graduates become inputs to society, as well as new inputs returning to higher education itself. It in turn requires inputs from society in the form of new students, financial support, personnel, and so on. The relationships with the larger society become the constraints within which higher education must operate.

The curriculum is an integral subsystem of higher education and grading, an integral part of the curriculum, serves a variety of functions both within and outside of higher education, as was seen above. For instance, employers and graduate schools look to grades as indicators of achievement and expect reports to be given in the form of grades. Grades also, as has been seen, have important effects upon the teaching-learning process and upon the recipient of grades, the student. The close relationship between grades and the curriculum was likewise described earlier, when it was noted that the introduction of the elective system led to the general characteristics of the individual course exam and grade. In addition, as one facet of undergraduate higher education evolved, namely mass education, increased pressure developed to require standardizing measures of learning and grading to simplify their usage and handling.

The recognition of such interrelationships of system and subsystem as these formed the basis of the research cited above in Making the Grade by Becker, Geer, and Hughes, who analyzed student life as a social system and observed three primary subsystems—academic work, campus activities, and

\textsuperscript{363}Ibid.
personal relationships.\textsuperscript{364} Regulations concerning courses, grades, credits they saw as constraints or givens from the formal and academic subsystem which structures student living, and they contended that grades are not just a measure of achievement, but are rather an important part of the larger college system which helps to define the student's self-worth and the like.\textsuperscript{365}

In addition to the complex hierarchy of interacting parts which "system" implies, it also signifies plan, order and function. System and subsystem must be understood in terms of function, purpose, and objective. For example, in her study of early grading systems Smallwood declared:

In order to consider the methods of examining students and the means of evaluating these examinations and by implication those who took them, it is necessary to keep in mind a few very definite points. First, no procedures which became educational methods can be set off by themselves for consideration. For example, the development of examinations and grading systems can be considered only when recognition is given to their use in setting entrance requirements, in the awarding of degrees, or in connection with general university regulations, for the ends in view influenced the methods used.\textsuperscript{366}

The end of purpose, she concluded, is the ultimate criterion according to which system is rational, effective, and efficient. A definition of "system" includes such an understanding:

We use the word "system" to indicate nothing more or less than the identification of all parts, working independently and in interaction to accomplish previously specified objectives. In our parlance, then the System Approach is the application of tools and techniques of system analysis and system synthesis to assure that the

\textsuperscript{364}Becker, Geer, and Hughes, \textit{Making the Grade}, pp. 2-3.

\textsuperscript{365}Ibid., p. 8.

total system will do what it is supposed to do under the conditions under which it must operate.367

Conditions for a desirable instructional system were developed by Ikenberry: (1) Time restrictions should be abolished—instruction should continue until mastery is achieved. (2) The student should understand instructional objectives and how they relate to his immediate and long-term plans. (3) The objectives should be stated in unambiguous terms. (4) Student involvement in the learning process should be maximized. (5) Instruction should provide accurate, timely, and informative feedback to the student concerning his progress toward objectives. (6) Emphasis should be placed on positive reinforcement. (7) There should be an appropriate sequence of learning experiences, diagnosis of learner deficiencies, and corresponding adjustment of the instruction sequence. (8) Timely and reliable information on individual student progress, and adaptations for the individual learner are necessary. (9) The total learning environment of the student should be considered in developing instructional goals and processes. (10) Different instructional systems should be available so the student can choose according to his own learning style.368

Several steps to instruction were also developed by Tyler, in the context of a learning system:

The first of these is to decide what ends to seek, that is, what objectives to aim at or, stated more precisely, what changes in students' behavior to try to bring about. The second step is to determine what content and learning experiences can be used that are likely to attain these ends, these changes in student behavior. The third step is to determine an effective organization


of these learning experiences so that their cumulative effect will be such as to bring about the desired behavior changes in an efficient fashion. Finally, the fourth step is to appraise the effects of the learning experiences to find out in what ways they have been effective and in what respects they have not produced the results desired.369

For purposes of this study, system is used to indicate an array of specific elements and activities organized in particular ways in order to obtain specified objectives. Results are judged in terms of desired objectives, and by means of various feedback functions, results can be compared to desired objectives, so that appropriate corrective actions may be taken when deviation is observed. It is also recognized that each subsystem can be a subunit of a larger system and in turn be composed of smaller subunits. The systems approach is considered an appropriate context for the consideration of grading.

Evaluation

When the function of the system is defined and the structure of the system is determined, then control or evaluation becomes possible. Newman and Summer pointed to three elements of the control function: (1) a standard that represents desired performance; (2) a comparison of actual results against the standards; (3) corrective action.370 Criteria or standards arise from the objectives and functions a system is to serve; the comparison of actual results to standards is properly called evaluation; and the corrective action can be called decision making, whereby activity is undertaken to alter inputs or the system itself so as to obtain the desired outputs.


Grades or the grading subsystem of the curriculum sub-system of higher education can be a means for evaluation. Johnson, Kast, and Rosenzweig noted that "Control can be added to an educational system by measuring the student's understanding of the course material through testing." They added: "As a corrective input the instructor might adjust his method of presenting the material, and additional work assignments, or take any other action which the feedback information seems to indicate."

One essential feature of the control process in a system is its feedback nature, which allows inputs or the system itself to be changed to achieve the desired output. The output in education is learning by the student. The desired output will in each case be determined by the function or functions of the system in question. A definition by Guba emphasized the purposive nature of evaluation: "...a process of providing and using information for making educational decisions." He continued on to suggest that if one can assume different kinds of decisions are to be made that there will be different kinds of evaluation.

Evaluation was also defined by Stufflebeam in relation to decision-making: "Generally, evaluation means the provision of information through formal means, such as criteria, measurement, and statistics to supply rational bases for making judgments which are inherent in decision situations."

371 Johnson, Kast, and Rosenzweig, Theory and Management, p. 81.
372 Ibid.
374 Ibid.
"evaluation is the science of providing information for decision-making." According to Stufflebeam, there were four major decision situations in education: (1) Planning decisions, which focus on improvements by specifying the domain, major goals, and specific objectives. (2) Programming decisions, which specify procedures, personnel, budget, and other resources for implementing planned activities. (3) Implementing decisions, which relate to directing programmed activities. (4) Recycling decisions, which concern terminating, continuing, or modifying activities.

Since evaluation relates to decision-making, there are also four kinds of evaluation. First, context evaluation defines the "environment where change is to occur, the environment's unmet needs, and the problems underlying those needs." Second, input evaluation is to "identify and assess relevant capabilities of the proposing agency, strategies which may be appropriate for meeting program goals and designs which may be appropriate for achieving objectives associated with each program goal." (Decisions resulting from input evaluation are in the form of procedures, materials, schedule, staff, and other resources).

Third, process evaluation provides "periodic feedback to project managers and others responsible for continuous control and refinement of plans and procedures. The overall strategy is to identify and monitor, on a continuous basis, the potential sources of failure in a project." Information from process evaluation may be used to modify the process and to interpret the final outcomes.

376 Ibid.
377 Ibid., p. 30.
378 Ibid., p. 32.
379 Ibid., p. 33.
380 Ibid., p. 35.
Fourth, "Product evaluation is used to determine the effectiveness of the project after it has run full cycle." 381 The method is to operationally define and measure criteria associated with the objectives of the activity, to compare these measurements with predetermined absolute or relative standards, and to make rational interpretations of the outcomes using the recorded context, input, and process information. 382 Information here is used to decide to continue, terminate, modify or refocus a change activity.

These decision and evaluation-shapers were developed by Stufflebeam to apply to complex educational projects, but the application to teaching and grading is direct. Context evaluation inputs come from the curriculum, student needs, and the philosophy of the college, and they culminate with the objectives of the particular course. Input evaluation considers the capabilities of the students, course design, teaching method and strategies. Process evaluation provides feedback to the instructor and students and may take such forms as personal observation by the instructor, tests, student comments. Product evaluation relates to the final judgment by the instructor of the achievement of course objectives by the students.

A related approach to the subject was formulated by Stake, who proposed a 3x4 matrix of evaluation and suggested three types of evaluation. Antecedent evaluation, according to Stake, concerned any condition existing before teaching and learning which could affect the outcomes. In transaction evaluation he included the student-teacher interactions, teaching materials, assignments, and all other factors which related to the process of education. Outcome evaluation, last on his list, comprised all those abilities, achievements, attitudes, and aspirations of the student which resulted from an educational experience. 383 Evaluation, he suggested, occurred in the specification of intents for each level, in observation of what actually took place (as distinguished from what was intended), in formulation of

381 Ibid., p. 36.
382 Ibid., pp. 36-37.
absolute or relative standards, and finally in the making of judgments. Stake included both the context and input evaluation of Stufflebeam in the category of antecedent evaluation, and his application to the course and grading is the same.

Evaluation was also seen in a broad framework by Ralph Tyler. According to Tyler, it is an appraisal process which involved acceptance of specific values and the use of a variety of instruments of observation, including measurement, as the bases of making value judgments. His description of the process of instruction as a system, which was quoted above, closely parallels Stufflebeam and Stake, though expressed in different terms. He explained his fourth step or evaluation:

Obviously, this fourth step is educational measurement, or achievement testing. It is an essential part of instruction because without appraisal of the results being attained, the instructor has no adequate way of checking the validity of his judgments regarding the values of particular learning experiences and the effectiveness of their organization in attaining the ends of education.384

Intrinsic and pay-off evaluation have both been delineated by Scriven. The former, he insisted, is concerned with such things as content, goals, grading procedures, and teachers' attitudes, and is inclusive of context, input, and process evaluation. According to Scriven, both forms are necessary, so that if results are deficient the cause can be determined, and so that intermediate goals do not replace final goals.385

The concepts of "formative" and "summative" evaluation have become a part of standard phraseology in recent

---

years and have been defined by a number of writers. A fairly complete explanation was provided by Airasian:

In mastery learning, the primary purpose of summative evaluation is to grade students according to their achievement of the course aims. Summative evaluations are in the realest sense "final" and grades assigned on their basis are likely to follow the student throughout his scholastic career. Summative examinations occur infrequently, typically covering relatively large blocks of instructional materials.386 An accurate means of ranking students with respect to their mastery of the overall course objectives is sought. Those who attain the pre-defined mastery level receive an A grade or some other suitable indication of mastery performance. Those who fail to attain the mastery level receive appropriately lower grades.387

His definition of formative evaluation is as follows:

Basically, formative evaluation seeks to identify learning weaknesses prior to the completion of instruction on a course segment—a unit, a chapter, or a lesson. The aim is to foster learning mastery by providing data which can direct subsequent or corrective teaching and learning.388 In keeping with its aim, formative evaluation should occur frequently during instruction. It should strive to identify unmastered learning areas early enough to permit their correction before the grading evaluation. .... Students need to be informed of their non-mastery at a time when they can, if they choose or if the instructional mode permits, correct their errors.389


387Ibid., p. 79.

388Ibid.

389Ibid., p. 80.
While Airasian is considering evaluation in the context of mastery learning, the concepts apply in the same way in other situations. In Stufflebeam's terminology, formative and summative evaluation are process and product evaluation.

Carroll defined formative and summative evaluation in the same manner, stressing the importance of formative evaluation consisting of tests or other indicators of learning given at frequent intervals for the purposes of diagnosis and remediation. The student, he said, needs clear information on what he has learned and still needs to learn.390

Block, in defining these types of evaluation, commented that summative evaluation cannot guide the teaching-learning process, but that formative evaluation provided immediate and continuous information regarding a student's progress during instruction. Formative evaluation should consist of brief diagnostic tests which are integral parts of teaching and learning. Block suggested correcting but not "grading" formative evaluations.391

Bloom distinguished these types of evaluation by their purpose, by the portion of the course covered, and by the level of generalization. He advanced the theory that formative evaluation helped learner and teacher to focus upon the particular learning necessary for mastery, to determine the degree of mastery of a given learning task, and to pinpoint the part of the task not mastered and determine why. Since such evaluation is diagnostic, it should not be graded, Bloom decided. Summative evaluation, as he defined it, attempted to assess the degree in which larger outcomes have been attained over the entire course or some substantial part of it, further purposes being to provide the basis for grades and reports to parents or administrators. This evaluation occurred after learning, Bloom noted, pointing out

390 John B. Carroll, "Problems of Measurement Related to the Concept of Learning for Mastery," in Mastery Learning, ed. by James H. Block, pp. 36-37.

391 James H. Block, "Operating Procedures for Mastery Learning," in Mastery Learning, ed. by James H. Block, pp. 64-76.
some danger in attempting to combine both forms of evaluation on the same test. 392

"Evaluation," "grading," "marking," "measurement," and related terms are used in different ways by different writers, and understandably a different usage sometimes results in disagreement over the desirability of and need for these procedures. For example, Wolff distinguished three types of grading. Criticism he defined as desirable feedback on performance; evaluation as the measurement of performance against an objective standard of excellence (which may be as simple as pass-fail or may encompass finer gradations). Ranking, according to Wolff, is the relative comparison of students for the purpose of determining a rank extrinsic and harmful to education, in that it is designed for selection purposes and justified on economic grounds, and he made the concept of evaluation include both comparison to an objective standard and the assignment of a grade. He believed that such evaluation was extrinsic and was made simply for the sake of service of graduate admissions. 393

In previous definitions, evaluation does imply a comparison to standards, but it is difficult to understand how the "criticism" of which Wolff approves does not.

His presentation contended that product or summative evaluation may but does not necessarily have to be expressed in the form of a grade, nor does it have to be reported or recorded. It seems correct to conclude, therefore, that Wolff was objecting primarily to the assignment, recording, and reporting of grades and to the comparative ranking of students which this makes possible, rather than to the concept of evaluation itself. In two articles, Manello, a strong critic of grades, argued for precisely this approach, namely, detailed feedback or evaluation for students, but no letter grades. 394


393 Wolff, "Ideal of the University."

In two articles concerned with the measurement of educational achievement and grading, Robert Ebel also distinguished between three related concepts. He defined measurement and evaluation as:

. . . a quantitative description of how much a student has achieved in relation to the achievement of his peers. A measurement is objective, impersonal, and can be quite precisely defined in operational terms. An evaluation, on the other hand, is a qualitative judgment of how good or how satisfactory the student's performance has been. Evaluations tend to be subjective, quite highly personal, and difficult to define precisely.

Marks were viewed as symbols used to summarize the results of measurement. Evaluation to Ebel included value judgments about a student's worth, motivation, and industry, in addition to his achievement. He believed, therefore, that evaluation could not be valid or reliable since it was so subjective, and that even if this were not so, achievement alone should still be the basis for marks. Obviously, evaluation as defined by Ebel is quite different from evaluation according to Stufflebeam and others. Contrary to Ebel's position, product or summative evaluation ought not be subjective and unreliable and should be based on achievement alone. Formative or process evaluation may be concerned with these other factors insofar as they facilitate or hinder achievement of desired objectives, but in this sense Ebel's criticisms would not apply.

A number of other writers have noted that evaluation was if properly understood an integral part of instruction and learning. Dressel and Saupe stressed that

395Ebel, Measuring Educational Achievement. Ebel, "Basic Considerations."

396Ebel, "Basic Considerations," p. 8.


399Joe L. Saupe, "Learning and Evaluation Processes,"
evaluation led to a clear statement and understanding of objectives which should be helpful to the student and concluded that careful evaluation avoided the situation wherein explicit (and usually higher-order objectives) are replaced by the actual or implicit objectives which appeared on tests and other measures of achievement and were more easily measured.

The problem of terminology and interpretation of practice into terminology has been and will continue to be a problem. Pace, for example, defined evaluation as a broader concept than "measurement," with evaluation including emphasis on the specification of objectives and their attainment. Cook, on the other hand, used measurement in the same broad way. Apparently they agree in substance but not in terminology.

In an article titled "Grading Evaluation," Erickson declared grades to be administrative shorthand which facilitates the process of ranking and classifying students, but he maintained that they do not evaluate specific educational achievement. Evaluation, he visualized, as taking many forms, such as student-instructor conferences, comments on exams and papers, and so on, defining evaluation as "... the process by which the student is informed of how well he is achieving the goals the teacher has set for the class ..." While Erickson's understanding of evaluation is sound, and while his description of grading based on ranking may be frequently true in practice, it is not clear that grading could be based on the results of evaluation and the achievement of goals. An argument similar to Erickson's was also presented by Pemberton.

---


402 Ibid.

403 Pemberton, "Grade Point Average," p. 9.
Summary

Higher education is a subsystem of the social system and receives significant inputs from this larger system. It in return expects certain outputs, typically in the form of new knowledge and in graduates with certain talents and skills, certain values, and the ability to think and judge, graduates who will contribute to society. As a sign of achievement, society looks to such credentials as degrees and grades for general certification and for certification in specific areas.

The colleges and universities which comprise higher education are an interrelated group exchanging students and graduates and generally, with some latitude for individual variation, adhere to certain generally prescribed formats and conventions to facilitate exchange and understanding. Curricular programs have been organized to fulfill functions of the institutions, and are usually organized into arrays and patterns of discrete subunits with levels and sequences according to prerequisites; these are called courses. Upon the successful completion of a subunit or course, a standardized symbol is recorded. Upon the completion of the entire array or pattern, a degree of certification is awarded which is accepted by society as indicating that achievement. Context and input evaluation are required in setting up curricular programs and courses for the achievement of institutional objectives. In the conducting of courses, process or formative evaluation is seen as necessary and desirable for effective attainment of objectives. At the end (and perhaps during) a course or a complete curriculum, a product or summative evaluation is called for to determine whether the desired objectives have finally been achieved. This complex process may be summarized in descriptive form or could be in the form of letter symbols as has been traditional.

This final step or "grading" is not, as some critics have noted, an essential part of summative, much less of formative evaluation. It does, however, have a significant impact on students for a variety of reasons, including its use by other systems with which higher education interacts. It should finally be noted that higher education is not completely free to decide to use or not use "grading" since the other systems impose constraints on higher education.
In the process of evaluation outlined above it was specified that reference must be made to norms in order to judge process and outcomes. Further, these norms were described as either normative or absolute, without further elaboration. Yet a central question in grading today is precisely whether absolute or relative norms should be used in assigning grades to students. As usual, this is not a new question. In reviewing the history of grading it was shown that the percentage system of grading was dropped, at least in part, because it was difficult to defend a particular grade as actually representing, for example, 77% of the absolute knowledge in a particular field. When letter grades were introduced they were frequently defined in normative terms.

Normative and Criterion Reference

Grading systems based on curving or on a normal distribution were considered in some detail above, when it was noted that a student's grade under this approach reflected achievement indirectly, insofar as the grade is based upon the achievement of the class as a whole. The effect of grading on a curve in continuing courses was seen and finally several methods for adjusting grades according to student ability were discussed.

One of the best explanations of norm-reference grading has been supplied by Robert Ebel, who believed that absolute standards are impossible to define objectively in most fields of study, since it is impossible to know the theoretical sum total of knowledge existing in a field of study, and also to know what proportion of that indefinite sum a student possesses. Ebel implied that there were no units to measure the total to be learned or the part the student does learn. Therefore, he said:

Grading on the curve, or more precisely, grading on the basis of observed differences in student achievement, is justified. In fact it is the only logically defensible and practically effective basis for grading.405

---

404 Ebel, "Basic Considerations," p. 5.

405 Ibid., p. 4.
He preferred to call this method "relative" grading, an approach which ideally should define grades on the basis of standard deviations from the mean rather than an ideal normal curve, so that if distribution of achievement is skewed, grades will also be skewed. He argued that all areas of human activity were for persons outstanding in relative and not absolute terms: "The winner in any race is determined on a relative basis." The competition implicit in this system was seen by Ebel as stimulating the maximum effort of each student.

Some adjustment for such differences in ability as were assumed to occur in small classes are allowed in Ebel's system: an experienced teacher, for example, could use norms generated by combining achievement in a succession of small classes, so long as the accumulated proportion of grades should approximate those called for by the normal curve. Such grading then would not be relative to the achievement of the individual class, but rather to the combined classes. Ebel also pointed out that the grades in a remedial or introductory course should not be lower than those in an upper level specialized course for high ability students, or vice versa:

... if the courses enrolling students at different levels ability are clearly identified, and if the grades are interpreted as indications of how well the student succeeded in tasks considered appropriate for his level of educational development, then differential grading standards are unnecessary and undesirable.

The reason he forwarded for this not being acceptable is that it introduced an attempt to measure in absolute terms what is better described and understood in relative terms. Again, according to Ebel, "The basic principle of relative grading is that a student's achievement should be graded by comparing it with the achievement of the student's peers." A music major consequently should be judged in relation to his or her peers in the course concerned, and not to more or less

---

406 Ibid., p. 6.
407 Ibid., p. 6.
408 Ibid., p. 7.
advanced students in other courses or to students in other fields who are pursuing music on an avocational basis. Finally, Ebel concluded:

If we reject the notion that grades are or can be absolute measures of achievement, and if we learn always to ask what they are relative to—what course work and what group of students—we should find fewer problems in grading, and should give more meaningful and reliable grades.409

Four comparative groups possible in normative reference grading have been outlined by Thorndike. Performance may be relative to a national group of a given grade or age; the comparative group could also be a total grade group in a smaller educational unit or the specific class in which a student finds himself—which is the most typical; finally, it could be the vague and ill-defined group which constitutes a particular instructor's concept of what is typical, based on his teaching experience and past classes.410

The normative approach has been criticized by many. Marshall stated his strong belief that ranking of students according to better and worse did not contribute to effective instruction.411 Back in 1933, Crooks noted that opposition to the use of the normal curve had arisen from various persons, for a number of reasons, two such reasons being that marks had no real meaning, and that achievement could not be expected to follow a normal curve.412 Gold noted that on this basis a class of students which on the average thoroughly mastered a particular course would receive the same average grade as another class which did not learn as much. Further, this policy, he thought, put students in competition with each other, rather than with standards of

409bid., p. 7.
410Thorndike, "Marking Systems."
411M. S. Marshall, "Your Grades Are?" College and University, XLIV (Winter, 1969), pp. 182-188.
Concerns about competition among students and similar problems with normative grading have also been discussed above.

Criterion reference grading or absolute grading, however, according to Glaser, assumed the notion of a continuum of knowledge acquisition, ranging from no proficiency to perfect proficiency. Objective or competency based curricula have a generally similar meaning: the student's performance is compared to a standard of behavior which defines the various points along the achievement continuum. Criterion referenced measurements give information about the degree of competence attained by individual students independent of reference to the performance of other students. Specifically, criterion reference tests find satisfactory items which all students may answer correctly, providing the item related to and accurately indicated achievement of a course objective. Such an item would normally be omitted from a norm referenced examination.

An interesting exchange between Lacey and Phillips on this matter was published in the AAUP Bulletin, Lacey arguing for a precise variability procedure in combining results of several tests based upon the value of a completely normative grade, and Phillips countering that grading should not be concerned with individual differences as such, but rather with student achievement.

According to Trow, an instructor should specify course objectives which relate closely to the course content and should determine, and explain to the student, a scale of

---

413 Stanford C. Ericksen, "Grades," Memo to the Faculty, No. 18 (Ann Arbor, Michigan: Center for Research on Learning and Teaching, the University of Michigan, 1966).


excellence against which his achievement will be evaluated. This "absolute" measure of achievement should then be used, rather than grading on a curve, and should include tasks which contribute to a distribution of scores. In this system, grades would be based on the absolute scale of excellence. Troyer likewise suggested that there be a set of detailed competencies for every course, which would include achievements appropriate to college, departmental, and specific course objectives. He realized that able students might attempt to demonstrate competencies before the course is completed, providing that participation in activities is not intrinsic to the objectives. In that event, the instructor would notify the registrar that the competencies have been achieved and that credit is to be granted.

Other writers have argued for this approach implicitly and explicitly, whatever their terminology, some within the traditional letter grade framework and others on a strict pass-fail basis.

---


418 Troyer, "Where Have Grades Gone?" pp. 542-555.


Pascal, "Methods of Grading."

Pascal, "Alternatives to Traditional Grading Procedures."
The criterion reference approach has been well summarized by Fraley:

There must be worthwhile objectives at all cognitive, affective, and psychomotor levels. Test items must be addressed directly to these objectives, weighted according to their relative importance, and tested in proportion to those weights. Performance levels must be specified for each objective and all learners must attain them, so that learner progress is contingent on learner attainment of the objectives.\(^{420}\)

Teaching for Mastery

Several notable articles have been contributed to the literature of teaching and grading by Benjamin Bloom, who specifically stated that almost all students, given sufficient time and individualized ways to learn, can master each learning task. In order for them to do so, Bloom believed that standardization of instruction and textbook must be reduced, and that the presentation, explanation, ordering of elements, instructional materials, time allowed, and so on, must be varied according to the needs of the individual student. In this system, "absolute" norms were suggested, which can be taken from achievement of previous classes, since they were difficult to compose in the abstract. In this way students were made to feel that they are judged on a performance basis, rather than on competition among themselves. Bloom specifically rejected the normal curve, since it predetermined many students to partial success or even failure. Only when students with varying learning needs, and with ability distributed normally, received identical instruction in a precisely defined time period, would results be normally distributed. Bloom believed that education should not be based on chance, which leads to the normal curve, but it should be designed to respond to the needs of

The same approach was followed by Chronister, and Mayo has also described the mastery approach, but with reference to secondary education.

Another writer who has written about teaching for mastery, James Block, stated the degree of learning was positively related to student perseverance, aptitude, quality of instruction, ability to understand instruction, and the time actually allowed in relation to the time needed. In mastery learning, Block included the following facets: (1) specific educational objectives; (2) well-defined learning units; (3) requirement of complete mastery before progress to next unit; (4) ungraded diagnostic tests after each unit, to provide feedback on student learning; (5) appropriate learning correctives based on diagnostic information, so students can complete the learning unit. According to Block, this method allowed 75% to 90% of students to achieve what formerly was attained by only 25%, a worthwhile result, he concluded, since under this method all students tend to be successful and learning is rewarding to the student.

In another article, Block suggested that when it was assumed that many students cannot learn well or some learn better than others, then the instructor may focus on sorting the more able from the less able rather than on seeking

---


achievement by all. Ohmer Milton recommended "PIP" instruction (personalized, individualized, process) and defined it in a manner similar to Block.

The effectiveness of at least one approach to teaching for mastery was reported by Whitehurst and Madigan who found that poorer students who, under the Keller Plan or PSI, were allowed retests to demonstrate their competence in specified objectives, did not learn less than their counterparts who demonstrated proficiency with few tests. On this basis larger numbers of high grades can be justified in such mastery designed courses.

By way of a footnote, a survey of graduate schools' attitudes was conducted by the graduate school of one institution considering mastery learning and wishing to use a grade of "M" to indicate successful completion of a course. While the author suggested the survey results showed that their students would have little difficulty in having their work accepted, the data of the study indicated that about half of the respondents did not endorse the approach.

The mastery approach does several important things with respect to instruction. It firmly involves criterion-referenced grading. It distinguishes between formative evaluation and summative evaluation, and uses both. It assumes that all students can succeed at a high level, and requires flexibility in instructional techniques and in time allowed, which is normally not the case.

---


426 Milton, Alternatives to the Traditional, pp. 66-89.


428 Robert Swanson, "Transferability of Graduate Work Graded According to Mastery: A Survey of Selected Graduate Faculty and Institutions" (Menomonie, Wisc.: Stout State University, Graduate School, June, 1970). ED 049 712.
A similar program has been described by John Roueche, who argued in favor of individualized and flexible methods of instruction with individual performance goals stated by behavioral objectives and without norm-referenced evaluation. He stated:

Once a student is admitted to a given program, he should be graded according to his ability. Each course must have stated behavioral objectives against which each individual's program can be measured. The time allowed to accomplish any set of objectives must be left open to allow for individual learning rates. The instructor may use a credit/no credit grading system or an A, B, C letter-grade system. In the first system, the learner either passes the final examination and receives credit for the course, or he receives nothing. If he does not pass the final examination, he continues working until he achieves the required course objectives. In the A, B, C system three performance levels are established. The student can work for any grade. If he tries for an A and fails, he can try again or accept a lower grade. If he fails to perform at the C level, he receives no credit.429

Competency-Based Instruction

A number of writers have discussed specific courses or programs which incorporate all or parts of criterion-reference, mastery teaching, competency-based instruction, and related concepts. The instructional plan at National College of Education was described by Troyer.430 No grades were given or GPA calculated; rather, for every course a detailed set of competencies was specified which related to college, departmental, and specific course objectives and the student's transcript recorded credit awarded.


430Troyer, "Grades Have Gone?" pp. 542-555.
Three writers have described in detail how the same approach was used in an undergraduate speech course, an undergraduate psychology course, and a graduate course respectively during the 1960's. Since then these approaches have received wide publicity through such instructional methods as the Keller Plan, competency-based teacher education, and comprehensive approaches to undergraduate curriculum such as those being developed at Alverno College.

At Alverno College, eight areas of competencies were defined: communications, analysis, problem-solving, value judgments and decision making, social interaction, individual and environmental relations, awareness and understanding of the world, fine arts and humanities. Six levels were then specified for each competency. Following this set up, upon the completion of a level of a competency a CLU or Competence Level Unit is assigned but not a grade. Forty CLU's are required, with at least four in each competency, and all six levels in at least one competency. CLU's may be earned by means of traditional courses, off-campus experiential learning, or individualized study. The college thus hoped to design alternative learning experiences. Assessment is a critical matter. Faculty are empowered to certify certain CLU in specific competencies as students complete course experiences and an Assessment Center grants CLU's for


433Taylor and others, "Effects of Contingent Grading."


achievement prior to attending college and for non-traditional learning.

A similar approach is being implemented at Mars Hill College. As a beginning assumption it was determined:

A curriculum designed around competences would consist of three basic elements: first, an overall statement of competences to be acquired for a successful completion of the program; second, sets of evaluative criteria for each competence which define the proficiency levels required for successful attainment; and third, sets of experiences designed to assist the student in attaining the required competences.\(^\text{435}\)

There it was assumed that the liberal arts graduate should have the ability to "(1) formulate and examine purposes; (2) design and act upon means of executing these purposes; (3) assess the consequences of such actions."\(^\text{436}\) In order to do this, several components were considered necessary, of which the one closest to general education was an epistemological structure for exploring the ways in which man creates meaning or purpose which structure is analytical to categorize basic areas or approaches to human knowledge according to their role in the formulation of meaning or purposes. Areas included were symbolic communication, science, aesthetics, personal knowledge, ethics, and synoptic knowledge.

The researchers then carefully defined each area in terms of objectives to be met:

This degree should be an institution's certification that the holder has realized a distinguishable constellation of academic attainments. This certification should be based primarily upon comprehensive assessments of a candidate's achievements, rather than on indices heavily dependent upon time-in-grade factors. The examining procedures should be valid means of


\(^{436}\)Ibid., p. 31.
assessing proficiency within the various domains of competency constituting an institution's requirements for the baccalaureate. This presupposes, of course, that explicit degree criteria have been systematically developed and publicly set forth.\footnote{J. W. Harris, "Baccalaureate Requirements: Attainments of Exposures?" \textit{Educational Record}, LIII (1972), pp. 60-61.}

Synthesis Between Criterion and Norm Referenced Grading

It is clear that two divergent philosophies of grading are theoretically represented in normative and criterion referenced grading. The first assumes that there is something absolute in the achievement of large groups, and that achievement in reference to the group is the most objective or absolute measurement. But proponents of the latter say that there is nothing absolute in the achievement of large groups, in particular where learners are treated individually, and that grades should be assigned in reference to carefully specified educational objectives.

In practice, however, the two approaches can become confused and less distinct. Ebel, an outspoken advocate for sophisticated normative reference grading, for example, has suggested that strict normative grading apply in remedial courses and honors courses, without adjustment for class ability, since an A or C grade in either course would accurately indicate how well the student performed in tasks appropriate to that course. This sort of rationale, it should be noted, clearly depends upon some of the arguments also used by criterion reference. Ebel also suggested that the instructor may modify grades in small classes—where ability may not be normally distributed—in accord with his "experience" of what other similar classes have achieved. This also applies in some degree to criterion reference.

In addition, the computation of the grade-point-average, and especially the ranking of students according to GPA, is clearly contrary to many of the arguments advanced by Ebel, since this in effect equates apples with oranges, or remedial courses with honors courses. Writing in another
context, Ebel stated that with normative standard scores one loses the meaningful relation between the score on the test and the character of the performance it is supposed to measure. To be meaningful, he believed, the standard score must be related to test content as well as to the scores of others. He concluded:

In this presentation our purpose has been to emphasize the need for and to demonstrate the possibility of test scores which report what the examinee can do. Content-meaning in test scores supplements but does not replace normative meaning.

Similarly, Bloom, an outspoken advocate of criterion reference grading, suggested that the setting of specific course objectives--and corresponding standards for particular grades--be done in the context of previous classes, since it was difficult to establish absolute norms in the abstract, a judgment not unlike that of Ebel.

It appears that some middleground between the two camps is possible. Normative referenced teaching at its worst may pit students against one another in a zero-sum game and allow the instructor to give his usual run of course lectures, design a few tests which will lead to high scores and low scores, submit an array of grades to the registrar, and be satisfied that he is completing another successful semester, albeit, without really facing the problem of what the real purposes of the course may have been and whether they were attained by students.

Criterion reference at its worst could lead to the setting of trivial goals, and result in insignificant student achievement, yet some form of "relative-absolute" grading might incorporate the best features of both. In working out such a form, it would have to be remembered that in essence, the standards of reference in the evaluation process must not be relative to the group itself, since the competitive and other negative aspects would apply. Standards,

---


439 Ibid., p. 25.
however, could be relative if they were based on other groups (classes) extrinsic to the class being graded, in which case they would be absolute in relationship to the class, since their own effort and achievement would affect the grades granted but not the standards for judging. In the terminology of criterion reference, objectives would be carefully specified so that these represent reasonable expectations of student achievement in the context of the typical students for whom the course is designed and the amount of credit awarded based on the instructor's experience with previous groups of students in the same course. If this level of achievement is not consistent with what is desired, either credit, prerequisites, or other factors should be adjusted so that expected levels of achievement might also be adjusted.

In any case, these standards of achievement would not be absolute in an ultimate sense, but would be absolute relative to the achievement of a particular class. Inputs such as time and method could be adjusted in order that more students might reach specified levels of achievement and receive appropriate grade recognition, which would be impossible by definition in norm reference grading.

Examinations

Since examinations are closely related to grading practices and to educational practice, it is not surprising that some attention has been given to them in grading literature. It was noted earlier that when the elective system was introduced, individual course examinations replaced the comprehensive examination system. The history of grading showed that originally most tests were essay in form; objective examinations were introduced in the 1930's. Through the decades, there were some objections to the entire system of courses, credits and marks, and to individual examinations as leading to segmentation, to discontinuity of the individual's knowledge, to the gearing instruction to the immediate, and to the memorization of mere facts.440 Proponents

440 Douglass, Teaching for Self-Education.

of the objective exams criticized essay exams as subjective, unreliable, limited in breadth and therefore subject to sampling error. The lack of time, they believed, would result in inadequate composition by the students, and to too cursory a correction of papers by faculty.  

In the 1930's, Jones, a critic of the "new" or "objective" examination, said most such exams concentrated on detailed factual items, trite memorized forms from textbooks, or on items vague and ambiguous. As problem solving and other such areas of study were not measured well by this sort of test, the student would soon learn to study according to the type of test used. Jones did admit that better preparation of such examinations might minimize these problems.

At least some recent articles or books suggested that the situation has not changed greatly. Hanchett, Hatch, and Woodring argued that objective tests promoted memorization and did not lead to higher level educational goals as did another type of examination. Marshall, a strong critic of traditional grading practice, also believed that objective testing was harmful in that it promoted memorization rather than understanding.

While both objective and essay examinations—as well as other forms of evaluation—have their merits and demerits; and though it is not the purpose of this paper to specifically discuss types of evaluation, it is important to note


442 Jones, Comprehensive Examinations.


444 Hatch, "The Examination."


446 Marshall, Teaching Without Grades.
that grades have important motivational effects upon students, and that, depending upon the method of evaluation, the educational outcome can vary considerably.

The advantages of an "outside" examiner, someone other than the instructor, has occasionally been a topic of grading literature. Earlier the role-conflict of an instructor in being both helper and judge was discussed. Outside exams have been proposed to solve this and other problems; another proposal, one that appears simplest, is that an outside examiner (rather than the instructor) be responsible for the exam and student grades in each course. According to Jones, this solution would have such advantages as providing for a less subjective evaluation; permitting the instructor to educate students toward an objective goal; awakening student perception of the instructor as a helper; encouraging more searching and analytical teaching; motivating students to be more serious about learning. Jones conceded that some will object that only the instructor of the class himself can know what is to be expected of the group, but he countered that objection by pointing out the possibility of there being an unconscious and detrimental tendency in an instructor who did his own testing to adapt his teaching to his form of exams.

The more radical proposal to limit testing in any particular course to formative type of evaluation and to leave summative evaluation to others, was forwarded when Jencks and Riesman argued that one institution may well provide education, such as a college or university, and another take care of certification of achievement. Jencks and Riesman, Academic Revolution, p. 62.

447 Axelrod, "What Do Grades Mean?" Alex Page, "To Grade or Retrograde?" College English, XXI (Jan., 1960), pp. 213-216.


and Pemberton\textsuperscript{452} likewise suggested such external examinations, using the Graduate Record Examinations, perhaps, or the College Level Examination Program.

But if the tests are conducted by an outside party, the college or university must give up its autonomy to some extent, since the nature of the tests would more or less control the nature and scope of instruction, and then grading in its present form would of necessity be discontinued. Hampshire College, following this thinking, has actually moved away from exams and grades for individual courses, to four examinations covering four areas of basic studies, an examination based on the student's program of concentration, and a final examination in a specialized program of advanced and independent study.\textsuperscript{453} Programs like this, however, have been instituted in only a few colleges and their acceptance is uncertain.

Grading Variances

One of the major arguments of opponents of traditional grading has been the tremendous variation in the assignment of grades which occurred no matter what criteria are used as the basis of comparison, a variation which cannot be explained away. This section, as an approach to that problem, will consider the variations which occur both in student ability and in grade distributions, and the tendency of grades to vary in time.

Ability Distributions

In assessing the appropriateness of grade distributions, the ability distribution of students must be taken into account. It is generally assumed that different colleges attract students of different abilities, whether by careful selection by the institution or by self-selection by students. All research confirms this assumption. Hoyt's study of 169 institutions revealed that the average

\textsuperscript{452}Pemberton, "The Grade Point Average," pp. 15-17.

A freshman ACT-Composite score ranged from a high of 98 percentile to a low of 10 percentile. Hood found the same range in 38 colleges and universities in Minnesota.

Comparable differences occurred in different programs and subunits of the same institution. Courses, too, may be marked by varying ACT percentiles: a report of randomly drawn small classes of 25 to 30 students produced a variety of ability distributions, a finding especially observable where the larger population was skewed. Since most small classes are not randomly drawn, the probability of such skewed ability distributions is much greater than usual. Special sections, where low and/or high ability students are removed from regular sections, also change the ability distribution but possibly not the average, except in the special sections.

These irregularities in distribution make questionable the application of the normal curve to grading, but variations in grade distributions must also be considered in relation to ability, lest it be argued that ability differences justify grading differences. This problem is considered below.

---


457 Greene and Hicks, "Normal Distributions?" pp. 296-302.

Variability of Grade Distributions

Variability Between Institutions

Four surveys of grade distributions, reported from 1939 to 1966, did not take into account ability or other factors. In each case they reported great variation in the percentages of A, B, C, D, and F grades assigned at different colleges. The range of high to low was most frequently about 20% and sometimes more.459

A second and more extensive group of studies examined grading distributions in the context of student ability. In one case the correlation between the average grade assigned by an institution and average student ability was only an extremely low .05.460 Another study developed a correction factor for predicting GPA from a student's ACT score according to the institution attended which indicated that a student's GPA might vary by well over 1.00, depending upon the institution.461 A third researcher developed an Academic Performance Index based on a survey of 34,000 graduates. When he divided colleges into four categories of selectiveness, he discovered that the meaning of a grade in constant terms varied to such an extent that an A- at a school in the lowest selective category represented the achievement of a C or C+ in the highest category.462


Perry, "Grades and Grading Systems?" pp. 159-165.


A second correlation study showed only a low positive correlation between student ability and the grades assigned by colleges, and the writers concluded that it was understandable to be skeptical about "absolute" standards of academic performance, there apparently being no standard level of required performance for a certain grade across colleges.463

Still another writer reported a wide range of average grades (1.9 to 2.8) noted in 38 Minnesota institutions and also different student ability levels, but found that the two were unrelated.464 Two further studies were able to establish the validity of an adjustment factor affecting the GPA of students transferring from one institution to another, which in effect compensated for the variation in grading standards between institutions.465

The conclusion to be drawn from these studies of grade distributions at different institutions, whether adjusted according to student ability or not, is that institutions use very different standards for judging and reporting student achievement, so a grade at one institution cannot be said to represent the same level of achievement as the identical grade at another institution. Certainly where any correlation of GPA and ability exists, it would explain only a tenth of the differences at best.

Except for the variation in standards, the process of assigning grades seems to be somewhat similar in most institutions. Studies determined that a general regression equation for predicting GPA did about as well as specific equations developed for a particular institution, when an adjustment had been made for varying standards by means of a


constant.\footnote{P. Hoyt, "Generalized Academic Prediction in Four-Year College," Personnel and Guidance Journal, XLVII (Oct., 1968), pp. 130-136.} There is some indication that institutions accepting students with a wider range of ability granted a wider range of grades. The correlation between the standard deviation of grades and ability is low but was positively correlated.\footnote{Baird and Fiester, "Grading Standards," p. 4.}

Variability Within Institutions

More important than variations between institutions—which to some extent can be taken into account by the reputation of the college and other more objective means—are the discrepancies within institutions themselves which render grading inconsistent and perhaps even arbitrary. These are of several sorts and are discussed below.

Institutional Grading Standards. A number of studies have been undertaken to consider the grading standards of institutions through time, as related to student ability. The results are clear-cut. Seven studies demonstrated that significantly higher ability levels in successive incoming freshman classes at a number of colleges and universities were associated with relatively constant grade distributions or averages,\footnote{Richard H. Miller, "Students Show a Preparation Increase but No Increase in Grades Was Shown," College and University, XLV (Fall, 1969), pp. 28-30.} yet presumably higher ability levels should

\footnote{Juola, "Illustrative Problems," pp. 29-33.}
\footnote{J. E. Bowers, "Test of Variation in Grading Standards," Educational and Psychological Measurement, XXVII (Summer, 1967), pp. 429-430.}
\footnote{Ohmer Milton, "Grades and Grading," Teaching-Learning Issues, No. 2 (Knoxville, Tenn.: Learning Resources Center, The University of Tennessee, 1966).}
\footnote{Wilson, "Increased Selectivity," pp. 46-53.}
result in higher achievement and, therefore, in higher grades. Baird and Fiester conducted the most exhaustive study in this area, discovering that colleges where incoming ability levels changed, the correlations between changes in ACT Composite Means and changes in college GPA means were positive and ranged from .11 to .32. Such low correlations, they concluded, indicated little tendency by college faculties to adapt grading distributions in the light of higher student ability and the greater academic achievement on the part of students which might be expected.469

While most studies cited focused upon the effects of increasing ability levels upon GPA, it is probably correct to conclude that when (or if) ability levels decline, average GPA also remains the same. Furthermore, if teaching methods, equipment, laboratories, and so on improved as ability levels rose, then the improvement in the actual performance level of students very likely would be even more marked and the adoption of stricter grading standards more pronounced.

The obvious conclusion to be drawn is that college faculties as a whole have certain preferred grading patterns relatively independent of achievement. This does not mean, however, that in the context of a particular class or department achievement does not affect assigned grades. Rather it indicates that certain relative norms are observed by institutions as a whole and that the average grade assigned is not based on absolute achievement but rather on adjustable scales, in consequence of which the grades of an institution tend to remain stable. Though it will be noted below that average grades have risen in recent years, this does not change the conclusion that grades are not directly connected with achievement levels. It indicates, rather, that faculties are adjusting their relative scales so that more higher grades are awarded.

Field of Study. A large number of studies considered the grade distribution of various fields of study with interesting results. (The amount of research done is probably accounted for by the readily available data on grades by fields or departments).

469 Baird and Fiester, "Grading Standards."
The first group of studies cited merely reported on the differences in grading patterns by fields, the duration of this particular interest being indicated by a study of grading patterns in elementary courses made at Harvard as early as 1911. When the range of grades was studied in approximately 20 departments, the researcher found that the percentages of A's ranged from 35% in Greek to 1% in English; B's from 32% in botany to 13% in English; C's from 52% in English to 21% in Greek; D's from 33% in Spanish to 10% in fine arts; F's from 21% in engineering to 1% in botany, and that the average GPA by department ranged from a high of 2.79 in Greek to a low of 1.54 in engineering. The author noted that while some might argue that higher grades are given based on greater student achievement in particular fields his analysis revealed that lower ability students were disproportionately enrolled in those fields. The discrepancy is even greater, therefore, than the face value of the differences would indicate.\textsuperscript{470}

In 1953, Morris developed a grading index at the University of Georgia by subtracting the percentage of D's and F's from A's and B's. The median of the index in the arts and science division was 11; in professional programs, 47; and for the university as a whole, 30.\textsuperscript{471} In 1960, when Weiss surveyed some 260,000 grades in six large Midwestern state universities, he found that grades in education were significantly higher than those in business, liberal arts, and science courses, even when controlled for level.\textsuperscript{472} At Iowa State Teachers College, a very detailed study of grading patterns also showed great differences in grades assigned, with the grade average in education high at 3.08, and the

\textsuperscript{470}Foster, College Curriculum.

\textsuperscript{471}V. C. Morris, "High Grades and Low Grades," College and University, XXVIII (1953), pp. 317-329.

average in humanities low at 2.05. Similar marked grade variations were reported by others.

A second group of studies attempted to analyze differences in grade distributions between fields in the context of student ability, one of the earlier and better studies being made at the University of Chicago by Reeves, in 1933. Finding that departmental grade averages ranged from 3.26 to 2.30, with ten of thirty-four departments significantly higher than the university average of 2.56 and seven lower, Reeves obtained or compiled an index of student ability for the students in each department, based on grades they had received in other departments. An index of grading standards was then determined for departments, based on a comparison of the actual mean of grades given in a department and the previous index based on student ability. This analysis showed that, after ability was controlled, students in an easy-grading department could obtain a GPA of approximately 3.25, while the same students, for relatively the same level of achievement, might earn 2.00 in a hard-grading department. Therefore the discrepancy in grading proves even greater in this instance if ability is taken into account.

473 H. M. Silvey (Director), Study of the Grades Assigned at Iowa State Teachers College During the 1958-1959 Academic Year, Bureau of Research and Examination Services (Cedar Falls, Iowa: Iowa State Teachers College, 1959).


475 Reeves, Peik, and Russell, Instructional Problems.

476 Ibid.
Similar results were found at St. Louis University,\(^77\) one report indicating that average GPA of departments and ability measures had a negative correlation of \(-0.35\) and that the particular department with the highest GPA had the students with the lowest average ability.\(^78\) Another study indicated a low positive correlation between GPA and ability,\(^79\) and several studies concluded that various departments graded higher or lower than expected in terms of student ability.\(^80\)

In general, it must be agreed not only that standards for judging achievement are not consistent between fields, but that they vary widely, a discrepancy which cannot be explained by differences in student ability levels since taking ability into account may actually accentuate the discrepancy.

Instructor Variability. It is a truism that individual variability is greater than group variability. Up to now the grading variability existing within rather large groups, in either a college or a university as a whole or in a field of study, has been considered, but a number of researchers have also attempted to consider the consistency of standards by individual instructors. The first group of such studies reviewed dealt with those multiple section courses which provide some obvious controls; the second featured instructor variability in general.


\(^78\) Pemberton, "The Grade Point Average."

\(^79\) Chase, "Academic Talent."

\(^80\) Davis, *Great Aspirations.*


Harvey, *Marking Practices.*
Instructor Variability in Multi-Sectioned Courses

The consistency or lack of it in faulty grading standards has been clearly indicated by a number of studies. At the University of Illinois it was shown that one-third of some thirty instructors in a particular subject graded significantly higher or lower than their peers and it was concluded:

Until such a time as contrary research indicates that instructors vary within chance limits, the common practice of estimating a student's worth upon the assumption that a grade has a fixed value is an administrative expedient without experimental foundation.\(^{481}\)

All studies in this field reported the same findings.\(^{482}\) Juola, for example, found a correlation of 0.03 between student ability and the mean grades assigned by the instructor. The same study reported that of three instructors, each teaching three sections of the same course, one graded consistently high (about .44 grade point), and the second consistently low (about 1.52). The third was "on target," but only because his grades were high in one section, low in another, and correct in the third.\(^{483}\)

Juola also illustrated the possible effects of a special remedial section on grade distributions. Establishing a special remedial section, he found, and removing students for it from a regular course while still applying to them the pattern of grade distributions used in the regular course,


\(^{482}\)Eldridge E. Scales, "Variability of Grading Practices Among Instructors of a Multiple-Section English Course," College and University, XXXIII (Spring, 1958), pp. 334-336.

\(^{483}\)Juola, "Illustrative Problems," pp. 29-33.
would result in large numbers of such students receiving lower grades. Remedial students, however, would receive correspondingly higher grades after such a move.

By implication, just the opposite might occur with the establishment of honors sections, yet, one study of 230 honors students at Indiana University showed they did do as well in honors sections as in other courses. Perhaps grade distributions were altered to take absolute levels of achievement into account, and if not, then other factors, such as motivation, would be required to explain the absence of differences in this situation.

Several writers have suggested methods for obtaining greater grading consistency in multiple section courses and have even succeeded in obtaining such results. The methods they proposed included the discussion of course objectives; the setting up of common procedures; the definition of course content and at least some common examinations. Where such a system is not adopted, however, and that has been the rule rather than the exception, the studies indicated clearly that little or no consistency in grading standards or grade distributions existed.

General Instructor Variability

As noted several times above, most problems involving grading have a long history. In 1928, for instance, Miller

484 Juola, "Illustrative Problems," p. 32.


E. E. Scales, "Effect of Instructor-Agreement on Evaluation Upon Assigned Grades in a Multiple Section Course," College and University, XXXVI (Winter, 1961), pp. 201-204.
noted a tremendous variation in grade distributions assigned by various professors in the same department, \(^{487}\) and later Melville, \(^{488}\) Harvey, \(^{489}\) and Willingham \(^{490}\) reported such differences. In 1960, Voeks noted that in one physical science department the range of percentages of A's and B's combined given by individual instructors varied from 33 to 71\% and D's and F's went from 8 to 42\%. Furthermore, in a second physical science department, instructors' ranges for A's and B's were 20 to 54\% and 9 to 43\% for D's and F's, while in a humanities department the ranges were 30 to 64\%, and 0 to 17\%, respectively. \(^{491}\) Kirby, working on the same problem, cited the following figures for average GPA's assigned by individual instructors: lower division courses ranged from a low of 1.82 to a high of 3.88, with the median of the lower third at 2.22, median of the middle third 2.44, and upper third 2.82. Upper division courses ranged from a low of 1.97 to a high of 3.71, with the median of the lower third at 2.42, the median of the middle third at 2.76, and that of the upper third, 3.20. \(^{492}\) Subtracting the percentage of D's and F's from the percentage of A's and B's, Morris found at the University of Georgia that the sum varied from minus 12\% to plus 74\% while the average was plus 30\%. \(^{493}\)


\(^{487}\) Miller, "College Marks."

\(^{488}\) Melville, "Academic Decision Making."

\(^{489}\) Harvey, Marking Practices.

\(^{490}\) Willingham, "Grading Variations," p. 368.


When a study by Reeves, cited earlier, concentrated on grade averages controlled for student ability, results showed great discrepancy in grading standards between departments, and further analysis of grade distributions in one department, English, indicated how individual discrepancies were lost in group averages. Unlike many other departments, this department was found on a whole to be grading appropriately, according to the ability level of the students, yet even there individual instructors varied greatly in their standards. Reeves drew the following conclusions, based on the discrepancies he observed at the departmental and individual instructor level:

These considerations lead to the grave question as to the wisdom of retaining any system of qualitative marking as a part of the academic machinery at the undergraduate level... When all of these qualifications of instructors' marks are taken into account, it becomes apparent that there is great hazard in using these marks for any important academic purpose, such as the award of honors or dismissal because of poor scholarship. This study leads indisputably to the position that the basing of academic rankings of students upon instructors' marks is unsound. Possibly grades in courses may serve well some other purposes, but their use for award of honors, for determination of eligibility for graduation, and for the elimination of undesirable students from the institution should be discontinued.494

Again it is clear that the standards and criteria used by individual instructors to assign grades vary greatly, and while such differences are not explainable by variations in student ability it is just as unlikely that they can be explained by other legitimate factors, such as a teacher's ability, which might lead to greater academic achievement in a particular class.

Levels of Instruction. Although the analysis of grade distributions by course level has not been the primary purpose of most studies, a number have provided such information

494Reeves, Peik, and Russell, Instructional Problems.
with unanimity of results. The normal increase has been .4 or .5 of a grade point from freshmen to seniors.495

Some have justified this rise on the basis of the attrition of poorer students, resulting in remaining students perhaps being of higher ability and one study confirmed this speculation. At Hampton Institute it was shown that the mean freshman GPA was 1.97; the sophomore GPA, 2.36; and the junior GPA, 2.46, means which were significantly different from each other. However, when students were omitted from the freshman and sophomore groups who were not included in the junior group, i.e., mean GPA's were based on the same students, the mean GPA's were 2.35, 2.48, and 2.46 respectively for the three classes.496

Another argument used to justify higher grades at upper levels is the greater number of courses in one's major taken at that level, and the consequent greater motivation and specific ability levels appropriate to such courses. More research is necessary to determine whether such rising grade levels are justified, but in any case, studies of grades should take these differences into consideration in interpreting grading data.

Other Factors Relating to Grading Variability

Several additional factors have been discussed in the literature of grading, among them the fact that student motivation, and therefore achievement and grades, may vary between

Harvey, Marking Practices.
Reeves, Peik, and Russell, Instructional Problems.
Foster, College Curriculum.
Silvey, Study of the Grades.

required and elective courses. Though no direct evidence has been presented to indicate different grading patterns in such courses, a possible explanation of the lower grades in required courses might be the absence of any need by instructors to promote enrollment via the prospect of high grades for students. Bass noted interactive effects between fields of study and course level wherein grade levels might not rise at upper levels in some departments. Grades in small classes were found to have higher mean GPA's than those in larger classes, and in the larger classes a higher percentage of marks in the middle ranges were assigned. The difference may not be independent of other factors; for example, lower level courses tended to be larger in size, and class size in some fields of study tended to be larger than in others. Aiken also commented on the lower mean GPA in larger classes.

Relative Stability of Grading Patterns

It should be noted, at least in passing, that the grading patterns discussed thus far are very stable from year to year; no studies have indicated any rapid shifts in grading patterns, and where grade distributions were examined, stability was always reported. The term relative stability is used in the sense that institutions, departments and instructors tend to maintain their relatively high or low grading patterns, even though the overall level of GPA may change, as is indicated in the next section.

Rising Tendency of the Grade-Point-Average

Occasional anecdotal evidence indicates that assigned grades in recent years are increasingly in the higher ranges.

499 Harvey, Marking Practices.
501 Ibid.
   Harvey, Marking Practices.
In a typical article in the Chronicle of Higher Education, the President of the University of Wisconsin reportedly asked a faculty committee to study grading practices, his reason being that the mean GPA of freshmen on the Madison campus had risen from 2.49 to 3.01, and that other averages had shown similar increases.502 A year later, another article in the same publication reported on the tendency of the average grade to increase at the colleges and universities of California. Faculty were reported as recognizing differing faculty practices to be a problem, while universally rejecting administrative interference as an intrusion on academic freedom.503 One survey, judging by responses from 168 institutions which showed that slightly over fifty percent of all students received A's or B's, concluded that the majority of students should not be classified as superior students as such grades would indicate.504 In 1973, another writer concluded that the widespread disenchantment among students and teachers with the letter-grade evaluation procedure led to an inflation in the number of higher grades out of proportion to any parallel increase in superior student achievement, indicating that grades and grading have been rejected as an effective academic and professional evaluation system.505

Another educator summarized his evaluation of present affairs:

A cynical account of general grading practice today would describe the C as an indication of attendance, the B as attendance with work done, and A as attendance with work done on time. So few grades below C exist that they are not worth talking about.506


505Battersby, Typical Folly.

Two systematic studies of this matter have been reported in the literature. When the grading practices of 435 colleges and universities were surveyed in 1971, the most striking result of the study was the fact that over a four year period, 91% of schools had increased GPA's, 1% showed no change, and 7% showed a decrease and that at the undergraduate level mean GPA increased from 2.40 to 2.56, and graduate GPA rose from 3.22 to 3.42.

The most recent report by Juola noted changes from 1960 to 1973. He contacted some 485 institutions and 197 responded, with usable data being provided by 134. Differences between the GPA's for successive years were computed for each college and this data was used to determine an annual rate of change for institutions. Not every institution provided information for each year, but all rates of change were positive and were as follows: 1960-1965--0.016; 1965-1968--0.026; 1968-1970--0.056; 1970-1972--0.053; and 1972-1973--0.028. Total increase in mean GPA during the time period was 0.404, or approximately one-half of one grade point. The author concluded that more complete sampling may result in a slightly greater or lesser increase, but the uniformity of increase in different types, size, and geographic institutions has definitely established a national tendency toward "grade inflation." He suggested that this phenomenon may be a reaction to the too harsh grading during the "post-sputnik" period; to the high standards used previously to control burgeoning enrollments, a trend now totally reversed; to the introduction of pass-fail grading; to difficulty instructors find in grading or in their willingness to grade less well structured curricula; to faculty avoidance of student hassle resulting from the assignment of low grades; or to other factors.

It was concluded earlier that grading standards and overall grading pattern of an institution are not directly related to the average ability of the student body, and that changes in average academic ability do not lead to changes

507 Burwen, "Institutional Research Notes."

in grading patterns. So while it does appear that habitual grading patterns are changing, and changing in a strong upward direction, it is improbable that the reason is higher student achievement. Though this rise is not considered at length in the literature of grading, it is nevertheless one of the most significant changes in the area of grading.

Some critics, as we have seen, have called for the revolutionary step of abolishing grades, in light of all the problems of grading, but with the exception of a very few and scattered institutions, traditional grading reigns supreme. The rise in GPA's, however, to the extent that it is not related to achievement and does not differentiate levels of achievement, will destroy the credibility of grades. If the trend continues, grades will become invalid for most of the purposes cited earlier, and whether the formality of assigning them is continued or not, they will not be used and the revolutionary suggestion of abolishing grades will have been implemented, albeit de facto.

The Carnegie Commission noted that instructors in some institutions have considerably reduced their expectations of students, no longer attempt differential grading, and have renounced standards in the name of responding to individual student needs. The Commission members stated their unequivocal opposition to this trend, specifically arguing for the maintenance of differential academic awards based on merit, (which in the case of students would be grades) and demanding that achievement, rather than years of study, should be the criterion for a degree. Society, they prophesied, would not obtain the maximum contribution from each individual according to his potential when all rewards were equal.509

Publicity concerning the problems resulting from grade inflation may well be moderating the trend, but very recent information is not available.

Precision in Grading

A topic engendering some discussion in the literature of grading has been the matter of precision or the fineness desirable in grading distinctions. Surveys of grading systems in use have indicated the general use of a five-point scale, on the basis of five letter grades, but show that numerical values and weights are also being assigned to grades, with averages being worked out to the hundredth decimal point.

A number of writers have argued for finer distinctions and pointed out the inappropriateness of calculating a very precise measure of grade averages to the hundredth decimal—which results in a scale of four hundred units—based on an initial scale of five units. Bassett noted this problem and suggested that course grades consist of numbers with two decimals, reasoning this would better reflect actual gradations. In a sense this would mark a return to the earlier percent system, but it need not be defined in terms of total knowledge in a field as was the former system.

Several articles have argued for a revision of the present system from a five unit scale to a 13 or 15 point scale by adding +s and -s and assigning corresponding numerical weights computing grade averages. Stroup twice stressed the inaccuracy of a five-unit grading scale. On the traditional scale, a student receiving grades of C+ (3 credit hours), D (5), B (3), C+ (5) would have an average of 1.88, he commented; taking into account +s and -s, however, he pointed out would raise the average to 2.04. A student's academic destiny, honors, dismissal, etc., often may depend upon GPA, and since personal judgment often was no longer depended upon in the making of such decisions, Stroup argued that the student was entitled to the most precisely derived


GPA that the institution can provide. At an extreme it can be noted that under the five-point system five grades of C+ are equivalent to five grades of C-, where a finer scale would lead to an average of 2.33 and 1.67 respectively or where the five-point scale would lead to a higher average for four C-'s and a B- than for five C+'s.

On the basis of probability distributions, Kirby determined that the error from using a five point scale would result in the following discrepancies for a thousand students: 4 - +.33, 21 - +.27, 62 - +.20, 123 - +.13, 580 - ±.07, 123 - -.13, 62 - -.20, 21 - -.27, 4 - -.33. 513 A very informative table was prepared by Ebel showing how the measured reliability of any factor relates inherent reliability according to the number of categories or units employed in a scale.

<table>
<thead>
<tr>
<th>Number of Categories</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inherent Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.95</td>
<td>.63</td>
<td>.85</td>
<td>.92</td>
<td>.94</td>
</tr>
<tr>
<td>.90</td>
<td>.60</td>
<td>.80</td>
<td>.87</td>
<td>.89</td>
</tr>
<tr>
<td>.80</td>
<td>.53</td>
<td>.71</td>
<td>.78</td>
<td>.79</td>
</tr>
<tr>
<td>.70</td>
<td>.47</td>
<td>.62</td>
<td>.68</td>
<td>.69</td>
</tr>
<tr>
<td>.50</td>
<td>.33</td>
<td>.45</td>
<td>.48</td>
<td>.49</td>
</tr>
</tbody>
</table>

He concluded that broad categories reduced errors since there are fewer possibilities for erroneous marks. (In fact all error could be eliminated by having only one category).


However, errors become proportionately more serious. The choice, therefore, is to some extent a matter of a larger number of fine errors or fewer larger errors. And where accurate finer measurements can be made, reliability or accuracy is reduced by using fewer but broader units. Ebel concluded that admitted errors in measuring student achievement did not justify using broader categories of grading:

But in most courses the basis for grading is, or can be made to be, accurate enough to justify the reporting of 10 to 15 levels of achievement. In the long run the more accurate the reports of achievement, the fairer they will be to the student and the more useful to those who must evaluate his accomplishments and his potentials.

Clark reported some decrease in reliability resulting from fewer units, but he felt that it was not serious. Grading systems with eight and five units were compared by determining the ratio of the standard deviation of weighted individual averages to the standard deviation of the combined grades of all individuals and by correlation of first and second semester grade averages. No comparable studies were reported in the literature, so no firm conclusion can be drawn from empirical study.

Three empirical studies of the effect on student GPA's of finer or more precise units have been reported. Stroup reviewed 250 transcripts of students with many +s and -s, with several semesters of work, and with borderline or probation status, to study the effects of counting the +s and -s which were not actually counted in the GPA by the university. On the basis of the random error resulting from gross units, he expected that the GPA of students would be hurt or helped.

---

514 Ebel, Measuring Educational Achievement.


in the same proportion by finer units. However, when the results of 610 semesters were considered, in 179 cases the GPA was raised, in 330 cases it was lowered, and in 101 cases it remained unchanged. Fifty-eight students who survived academically on the gross scale would have been unsuccessful on the finer scale, and only five would have been successful on the finer scale who failed on the gross scale. These results were unexpected and predicted little injustice to students by eliminating them from school, contrary to arguments by the author in other articles. The author explained the results by attributing the changes to a systematic sampling bias caused by the deletion of students dismissed from school because of the chance lowering of their GPA.517 This explanation, however, is logically incorrect, since such a bias would have exactly the opposite effect and tend to reduce the imbalance.

Two other studies reported results consistent with Stroup's in situations where sampling bias, as noted above, did not occur. Pettitt arranged to have the faculty of an entire college assign grades on a 45 unit scale, as well as on a traditional scale. Some 2,200 students in 32 departments were involved. When equivalent standards were employed, 36 students received honors who would not have done so under the new scale and 37 who would have received honors under the new did not do so under the old. Likewise, under the new scale 61 students would have been on probation, but were rated satisfactory under the old, while 35 probationary students on the old scale would have been satisfactory under the new. In general, obviously, more students were hurt than were helped by the more accurate scale. The classification of 169 of 2,200 students among honors, average, and probation was changed by the finer scale.518

When Philbrick had instructors of 42 students who were subject to academic disqualification compute their grades in the traditional way, and then on a finer scale

517F. Stroup, "Grouping Errors in the Grade-Point Average," Journal of Experimental Education, XXXIV (Spring, 1966), pp. 31-33.

using +s and -s, 73 of 268 grades were changed downward and 39 upward, so 11 students emerged with a higher GPA, 32 with lower, and 9 with GPA unchanged. The writer who admitted frankly that he was looking for factors which would help students stay in school, concluded that more precision in grading had the opposite effect from the one he sought.519

The tendency for more precise systems to lower student GPA's is inconsistent with Kirby's prediction based on probability, and with the expectation of Stroup. One plausible and simple explanation might follow from probable student behavior at the end of an academic term, which amounts to outwitting the statistical system of traditional grading. Where students have established a "solid" grade unlikely to be affected upward or downward into another grade level by term-end work, they are prone to spend little additional time on the class. On the other hand, in a course where a student's work is bordering between two grade levels, i.e., either a B- or C+, the student will expend great effort to maintain the higher grade level or to raise the current course average the slight amount necessary to achieve the next grade level. On the less precise scale it matters not at all that the "solid" B declines to a B-, but raising a C+ to a B- has a large impact. More grades are likely to be in the minus range than in the plus range, therefore, and the less precise system, which does not include either the plus or the minus, tends to aid students. The trade-off described above would not affect the GPA of a student under the more precise system. Students, if this reasoning is correct, are inclined to direct their study according to the statistical intricacies of the grading system.

Pettit found that 278 students preferred a 45-unit scale to a five-point scale; only 43 opposed that choice. Written comments indicated that the majority of students thought the popular system would reveal their performance more accurately and could reflect improvement within a grade level.520 When Goldstein compared student opinion on a


520Pettit and Crawford, "Refined Grading Scale."
three- and a five-point system, the five-point system rated higher in providing feedback and in facilitating entrance to graduate school, but the three-point scale, which was favored overall, was rated higher on student motivation, lessened anxiety, creativity, true learning, "general evaluation," and fairness.521

It should be understood that pass-fail systems, or variants, such as honors--pass--and fail, are basically two- or three-unit scales rather than the traditional five-unit--as has been noted by at least one writer.522 (The implications of pass-fail grading were discussed earlier). Teaf, in summarizing opinion of participants at a conference on grading, reported that finer grade distinctions put more pressure on students.523 While this is apparently true in comparing a two- or three-unit scale to a traditional five-unit scale, it should not be concluded that this relationship continues in a linear manner with progressively finer scales. As early as 1921 Wood, in reacting to the critics of the inaccuracy of traditional grading who suggested a pass-fail system, expressed his opposition on the ground that concealing inaccuracy by a gross classification is not a remedy, since achievement is not distributed into compartments, good and bad.524

In summary, the rationale for compressing all academic achievement into five categories, assigning numerical whole numbers from 0 to 4, and then computing GPA's to the hundredth decimal, which amounts to a scale of four hundred units, is at best dubious. Nor does the admitted--but relatively small--inaccuracy of a fifteen-point system necessarily cancel its desirability any more than a physical description of a person should use only whole feet, since it would be impossible to be sure of the exact number of inches. The greater


523Teaf, "What Price Grades?" pp. 100-103.

theoretical and empirical accuracy of a finer system should be considered for paradoxically, and contrary to student opinion, a finer system in some sense might mitigate some undesirable effects of grading on learning by allowing students to study as they should without the incentive of playing a statistical game with grades. In short, the merits of more precise grading systems deserve more thorough investigation than the literature of grading has yet provided.

Reliability of Grades

It has become a cliche among critics of grades to cite their unreliability. In 1915, Rugg noted their questionable reliability. In 1921, Wood concluded that the inaccuracy of college grades and their unreliability was a notorious fact with which no one disagreed. In 1933, in his review of grading, Crooks noted the questionable reliability of grades. Modern critics of the system include Goode, Fishman, and many others.

Two basic problems were delineated by Thorndike in regard to reliability. The first assumed homogeneity of the criterion measure and arose from errors of measurement in the criterion, errors which may be classified as random, unsystematic, and unpredictable. The second problem involved heterogeneity in the criterion measure: factors used to judge achievement or levels of achievement vary from course to course, institution to institution, and time to time.

528Goode, "Grading and Testing."
Tetlow discussed the multiplicity of grading procedures which lack standard interpretation of grading criteria and also involve grading differences which concern error in measurement. Inconsistencies or variations in measurement listed by Lavin include: differing test types, the use of term papers, class participation, relative weights of factors, varying standards, subjective interpersonal factors affecting grading, and different criteria used by different instructors. Ebel was also concerned with criterion problems since, he said, the meaning of grades is neither clearly defined nor observed, and the assignment of grades varies according to many factors.

Several of the topics discussed previously obviously relate closely to the reliability of grades. The various determinants, for example, which were shown to affect grades may be considered to indicate either heterogeneity of criterion measures or errors of criteria measurement. The variation in grades as discussed in another section would also fall under errors of measurement and would lower reliability. Precision in grading or the fineness of the grading scale also has effects on reliability: a five-point scale would result in a decrease in actual as compared to theoretical reliability on the order of .10. It was shown by Kirby that a combination of error sources does not cancel, but rather increases, the total error, though not on a strictly additive basis.

Among a number of studies which correlated the grades earned from one term to another to obtain some assessment of reliability, was one by Rogers which correlated each term standing with every other term standing for four college classes to determine that the average correlation was .66, ranging from a high of .80, between term one and two, and a low of .48, between terms one and eight. In general, Rogers learned that correlations decreased as the time interval between terms increased.

---

531 William L. Tetlow, Jr., "STAG."
532 Lavin, Academic Performance.
533 Ebel, Measuring Educational Achievement, pp. 401-402.
The same results were found by Ferguson and Crooks, with the highest correlation being .81 and the lowest .30; adjacent semesters in the study were correlated at approximately .67, and those three or more away were correlated at approximately .42. Clark, attacking the same problem, reported correlations between first and second semester freshman GPA's in the .70's. A 1931 study controlling for instructor differences, kind of courses taken, and other factors found an average correlation of .81 between close semesters. A similar study by Willingham in 1963, which made adjustments for interdepartmental grading variations, found that correlations with the control were .39 and unadjusted were .32.

However, the correlations of student GPA's from term to term cannot be simply equated with reliability since student motivation, study time, nature of courses, and many other factors may be different and alter student achievement, and the decreasing correlation between more distant semesters may be due in large part to these effects. In general, such studies do indicate a moderate reliability for grades; according to these correlations, GPA may be about 40 to 50 percent true variance.

Other research used other approaches and statistics in considering reliability, finding, for example, that the correlation of GPA and aptitude tends to decline each semester from .53 in a student's first semester to .12 in the eighth semester. Humphreys reported similar results.

---

537 Clark, "Grade Point Averages," pp. 428-430.
To the extent that ability is a true predictor of achievement it may be concluded that the reliability of grades, never high, progressively declines as the student advances. When Crawford correlated aptitude scores with faculty grades and with achievement tests covering the same subject areas as the faculty grades, correlations with achievement tests were higher than correlations with faculty grades, indicating the lower reliability of the latter.\textsuperscript{542}

In one research project, five instructors were teaching five sections of a psychology course with approximately 70 students in each section. A final grade was based on four in-term examinations which were departmentally constructed and common to all sections, and on a final exam designed to cover the whole course, with content selected by each instructor for his own section. The general reliability obtained for letter grades was .80, with individual instructors ranging from .76 to .87. The average reliability of the instructor-conducted final exam was .57, as compared to the average reliability of the departmentally-constructed examinations, which was .67.\textsuperscript{543} At Indiana University a correlation was computed between the two split-half freshman grade averages and then adjusted using the Spearman-Brown Prophecy Formula. The adjusted coefficient was .84.\textsuperscript{544}

While most research into grading reliability has relied upon the semester to semester correlation of student GPA's, studies using different approaches have tended to report about the same levels of reliability. It can therefore be concluded that GPA's do have moderate reliability, in spite of the many factors which serve to reduce reliability. It is also clear that reliability can be increased by controlling these extraneous factors.


Summary

The history of grading reveals the gradual development of grading systems similar to those in use at present. Numerical ratings replaced prose and other evaluations and they were in turn displaced by letters. Eventually a cycle was completed when letters were given numerical equivalents and student grade point averages were again averages computed to the hundredth decimal point. There were many arguments about what factors should be included in a grade, about the reliability and validity of grades, their positive and negative effects upon learning, the distributions of grades according to high and low, and the precision of grading. History also shows that such a modern innovation as pass-fail grading was first employed more than one hundred years ago.

A variety of grading systems have been used during the years by individual institutions. As the curricular system changed from a common curriculum to provide distribution requirements and electives the individual course grade became the norm. The dominant grading system is the A, B, C, D, F with the limited pass-fail option. This grading system and others which are significant as a result of use or theoretical importance are used in the study.

The literature does not show general agreement concerning the objectives of grading. Even the most widely accepted grading objectives are criticized. Motivation, for example, acknowledged by most persons as legitimate, is considered to be coercion of students by some. The objectives considered most significant and most widely mentioned are included in the study. The specific objectives and grading systems used in the study are presented in Chapter III.

The outcome is analyzed in terms of the suggested objectives of grading. Some objectives appeared to be more or less achieved and others less so. In many areas the literature or research is not sufficiently developed to draw certain conclusions. These findings in the literature are compared in Chapter V to student judgment concerning how well specific grading systems achieve selected functions.

The final part of the literature review explores a number of technical and theoretical grading concerns,
particularly in respect to systems and evaluation theory. The latter provides a theoretical framework for interpreting the results of the study and the literature and also provides a basis for recommendations.

The findings of Chapter II or the review of the literature are summarized in some detail in Chapter V.
CHAPTER III

EXPERIMENTAL PROCEDURE

Introduction and Background

The Problem

The problem addressed by the study was twofold. First, a comprehensive review of the literature was provided to make past findings readily available, to guide research and discussion on grading, and to provide some theoretical framework for the consideration of grades. This review of the literature considered actual and significant theoretical grading systems, examined the purposes of grades, the ways in which grades are actually determined, the results or outcomes of grading, and such technical/theoretical issues as the theoretical context of grading, the variability of grades, and the precision and reliability of grades. The analysis of actual and proposed grading systems, the objectives of grading, and grading outcomes—in particular as they related to student activity and attitudes—provide the specific context of the empirical study.

Second, the specific objectives of the empirical portion of the study are to determine the following from student opinion:

a. Relative importance or desirability of the proposed functions or decision-making objectives supposed to be served by grades.
b. Overall measure of the utility of various grading systems in attaining these functions.
c. Differences in judgment among selected groups of students.

It is expected that the results of the study will for the first time provide an analytical and objective report of
student consideration relating to the very purposes or justification of grading, and examine the potential of various grading systems in achieving these objectives or purposes according to student judgment. The analysis of differential responses by various student groups may provide some indication of the necessity for considering utilization of different grading systems for different groups of students, in contrast to utilizing one general system. Taking into account the response of students regarding specific objectives and grading systems is necessary in order to avoid designing a theoretically elegant grading system which pleases faculty and administrators, but a system which may fail to recognize student behavior and attitudes which can render the entire effort ineffective or even dysfunctional.

Hypotheses

The study was designed to test two sets of hypotheses. Concerning grading objectives:

1. That grading objectives are most accepted by majors in professional areas and least accepted by underclassmen.

2. That certain objectives, such as selection functions and mark of recognition for achievement, will receive more acceptance than other objectives, such as means of rewarding desired behavior and providing motivation.

3. That certain objectives will be more acceptable to one group of students than to the others, specifically:
   a. Underclassmen will consider motivation and mark of recognition particularly undesirable as grading objectives.
   b. Students in professional areas will consider selection for graduate school and employer, and means of rewarding desired behavior, particularly desirable as grading objectives.

Concerning grading systems:

1. That the more traditional grading systems will receive the greater acceptance.
2. That grading systems will be more accepted by students in professional fields and less accepted by underclassmen.

3. That more traditional grading systems will be most accepted by students in professional areas, and that less traditional grading systems will be most accepted by underclassmen.

The grade point perspective as defined by Becker, Gear, and Hughes was explored in some detail in Chapter II. The socialization process of the student in college teaches him various sorts of grade-oriented behavior and shapes his attitudes and values. It was expected, therefore, that as students are socialized into the system, they would tend to be more accepting of grading objectives in general. If the wish to abolish grades altogether, as reported by a large number of entering freshmen and noted in Chapter II, is associated with negative attitudes toward the objectives of purposes of grading, then the underclassmen should tend to rate objectives low. Professional students were expected to rate the objectives of grading higher, because they are taught by more conservative and traditional elements of faculty; they are more oriented to conformity required in their professional roles; they are likely to be more grade-oriented, as a result of their need to obtain a job or to be admitted to graduate school.

The selection functions, which provide the basic necessity for grading, were expected to rate high, in that students do understand the impact of grades in this context and direct their behavior accordingly. Mark of recognition for achievement was expected to rate highly, in that the review of student attitudes showed that they expect an equitable system in which grades accurately indicate achievement. On the other hand, student behavior, judged by such things as attendance, attitude, and appearance, is not related to academic achievement, so it is to be expected that students would tend to resist the notion that grades should be used to control such areas, which are in the student's private domain.

Finally, assuming that freshmen are more resistant to the entire concept of grading, it would not be surprising if they would be particularly opposed to the grading objective of motivation, directly associated as it is with
various grading dysfunctions, such as conformity, pressure, competition, and recognition of achievement provided in the form of grades.

The objectives most closely associated with the orientation of professional students are based on selection, and these objectives were therefore expected to be rated high. Furthermore, since such students were more oriented to the requirement or professional codes and mores introduced in classwork, less objection to the objective of rewarding desired behavior was expected from them.

In regard to grading systems, the socialization of students into the grade point perspective was expected to lead to greater knowledge of and acceptance of grading systems which did not depart radically from traditional grading. Professional students were expected to be most accepting of grading systems in general, and of traditional systems in particular, because of their professional orientation and socialization into the system. Underclassmen were expected to react in the opposite direction.

In general, the results of the study were expected to indicate how students as a whole judge both the objectives of grading and the success of various grading systems in achieving these objectives, as well as show the possible diversity in reaction by distinct groups of students. (The latter area has not been reported on in the literature).

An adjunct study was conducted on the basis of information collected concerning students' grade-point-average. The relationship between GPA and student rating of grading objectives and grading systems was explored. It was hypothesized that students at different grade-point-averages would rank grading objectives and systems differently. This analysis was not planned as part of the original study.

Research Design

Objectives and Grading Systems

A number of rationales have been proposed concerning the purposes or objectives of grading. Subsequently, a number of grading systems and variations have been proposed and used, based on the different objectives. From a review of
the literature, seven objectives were chosen to be included in the study, the first three of which have an intrinsic relationship to the teaching-learning process, while the last four serve extrinsic purposes. A full discussion of the purposes or objectives is presented in Chapter II. The appropriateness of the objectives was reviewed and approved by members of the dissertation committee. The objectives are:

1. Feedback, teaching tool for instructor.
2. Feedback, learning tool for student.
3. Motivation for student.
4. Means of rewarding behavior desired by instructor (other than academic achievement—such as attendance, attitudes, etc.).
5. Mark of recognition for competencies or knowledge achieved.
6. Selection/screening tool for college and graduate school. Basis for selection, awarding honors, retention in college, etc.
7. Selection tool for employer.

Of the many theoretical and actual grading systems, six were chosen to be used in the study, choice being determined by their actual usage or their relative importance in the literature of grading, or by the fact that they represent a basis from which several variations have been proposed. The selection of grading systems used was reviewed and approved by members of the dissertation committee. Grading systems included are:

1. A, B, C, D, F with pass-fail option in one course per term, restricted to elective courses (4 point system).

2. A, B, C (grades of D and F are deleted from records and are not counted in grade average, no credit is earned)—grade-point-average is computed from A's, B's, and C's only, courses in which the student does not earn C or above can be repeated later for credit.
3. Descriptive grading. A brief, one paragraph evaluation by the instructor, such as: "Works hard, masters basic factual information. Shows some indications of creativity. Has difficulty in synthesizing and applying course material." The student's academic record is composed of the comments of instructors for all courses taken. No grade point average, etc., is computed; grades as such are not assigned.

4. A+, A, A-, B+, B, B-, C+, etc., (15 point system, +'s and -'s are counted in the grade-point-average).

5. Pass-fail system of grading for all courses (grade-point-average, etc., are not computed).

6. A, B, C, D, F (4 point system).

Questionnaire

A two-part questionnaire was developed to elicit and record student judgment of grading objectives and grading systems. A seven-unit Likert-type scale was used, ranging from 0 to 6. Zero indicated no desirability in regard to goals or no effectiveness in regard to grading systems. Six indicated high desirability or high effectiveness.

Four forms of the questionnaire (A, B, C, D) were prepared. In each case the sequence of objectives and grading systems presented to the student was randomly determined in order to prevent any sequential presentation effects from influencing the results.

The questionnaire was designed to be self-administered, with all necessary explanation and directions provided. Part I of the questionnaire called for a response to grading objectives and Part II to grading systems. An introduction to the questionnaire provided some explanation and asked the student to indicate class rank, estimated grade-point-average, sex, and course number where appropriate. A sample questionnaire is presented in Appendix A.
Procedures for Computation of Rank Scores

In order to study the hypotheses, student judgment was used as a basis for (1) measuring the relative desirability of objectives or goals to be served by grading; (2) determining the effectiveness of several kinds of grading systems in achieving these objectives; (3) obtaining an overall rating--utility scores--of the grading systems from (1) and (2). An indirect rating was desired to minimize a purely affective response to a particular grading system. Step two was an intermediate step, included for the purpose of obtaining an indirect rating for grading systems, information not used as such to test hypotheses.

Accordingly, each subject rated seven objectives of grading on a 0 to 6 scale. They then rated the effectiveness of particular grading systems in achieving each of the seven objectives, also on a 0 to 6 scale. A composite score was then obtained from the product of the rating of a particular objective and of the rating of effectiveness of the grading system in achieving that objective. For example, if a student on a 0 to 6 scale rated the seven objectives 4, 6, 1, 3, 2, 4, and 5 respectively, and rated the effectiveness of a particular grading system in achieving those objectives at 0, 5, 5, 2, 3, 4, and 4 respectively, then the composite scores for a particular grading system, defined as the "utility score," was taken as a measure of the overall rating; in this example it would be 83.

Subjects

Three groups of students were identified in 1972-1973 as reference groups for the study: underclass students who entered higher education recently without a high degree of socialization and without a strong or permanent commitment to a particular field of study; upperclass students with an arts and science orientation, that is, without a specific professional or vocational orientation related to their field of study; and upperclass students with a specific professional or vocational orientation.

All subjects were from the student body at Ohio State University and participated on a voluntary basis.
Underclassman subjects were freshman students enrolled in the University College. Upperclassmen included only junior and senior students majoring in appropriate arts and science or in professional major fields of study.

Subjects were obtained in three ways. After selecting upper level art and science and professional courses, permission of the instructor to administer the questionnaire was obtained. Professional courses included agronomy, animal science, business, education, and journalism. Arts and science courses included art, English, geography, geology, and history. The questionnaire was administered either by the instructor or experimenter. In either case it was explained that (1) the questionnaire was part of a study of student attitudes toward grading; (2) their cooperation would be appreciated; (3) participation was voluntary; (4) the instructions in the questionnaire should be read carefully and every question should be answered.

The questionnaire was distributed to underclassmen at an orientation meeting and the same instructions were given. Permission of the instructors and of the chief administrator of the University College was obtained through personal contact, after explaining the nature and purpose of the study. Although student samples were not randomly selected, there is no reason to believe that results were affected by the selection procedure.

Only those subjects who accurately completed the questionnaire were considered for inclusion in the study, so questionnaires were deleted from consideration when any answer was left blank. One hundred students were finally selected for each group, that number having been determined as appropriate for the desired sensitivity on the basis of the pilot study described in the next section. Where the total of valid questionnaires exceeded the desired one hundred, excess questionnaires were randomly eliminated.

Pilot Study

A pilot study was conducted to test the quality of the questionnaire instrument, the time required for completion of the questionnaire, and the number of subjects required to obtain sufficient power in the analysis of variance statistic to detect significant differences of a predetermined magnitude between mean scores for treatment groups.
Procedures

The questionnaire was administered to 28 female students in a general psychology course at the College of Mount St. Joseph. Completion of the questionnaire required ten minutes, which was judged a feasible time period for administration in classes at Ohio State. No difficulties either in administration or in understanding on the part of the subjects appeared.

Estimation of Required Sample Size

Error variances were calculated in the pilot study to permit estimation of the sample size needed in the actual study. The desired sample size was determined by using $\Phi$, a formula described by Winer.1 For the analysis of variance concerning grading goals the difference considered significant was 0.67. For the utility of grading systems, this difference was considered to be 10.00. With 21 and 18 treatments respectively, average standard deviations for the pilot study of 3.43 and 46.65, with set at .05, it was determined that the $\Phi$ values were .36 and .21, and that 30 subjects per treatment were required in the former case and 80 in the latter. As a result, the desired sample size was set at 100, to allow for possible variation in the value of for the actual samples. Means and standard deviations for grading goals and grading systems are presented in Tables 15 and 16 of Appendix A.

Data Tabulation

The student responses on the questionnaire were coded on IBM answer sheets according to a key. The form of each questionnaire was coded (A, B, C, D) and the key enabled all data to be recast into the format of Form A. The data from the IBM answer sheets was converted to IBM data cards by machine. A printout of the data from the IBM cards was reviewed to insure the accuracy of the cards and of the conversion. Where incomplete data occurred in regard to grading

---

goals or systems, the subject was omitted. Composite and overall scores for grading systems were computed as described in a preceding section.

**Data Analysis**

Hypotheses were tested, using a fixed, two-factor analyses of variance with repeated measures on the first (student groups) factor. A 3x7 (student groups by grading objectives) design with an unweighted means solution was used for hypotheses concerning grading objectives, and a 3x6 (student groups by grading systems) design for grading systems. Where a significant interaction was obtained, a further analysis of variance for simple main effects as defined and as described by Winer was conducted. Where significant differences were obtained, the results were further analyzed, using the Newman-Keuls Procedure, to determine the source of the difference. Design layouts and tables of group means are included in Appendix B. The analysis of variance tables and Newman-Keuls tables are presented in Chapter IV.

In the adjunct study of the relationship of GPA to student ratings of grading objectives and systems, separate Chi Square analyses were run to study the relationships for each objective and each grading system. The chi square table for grading objectives was based on half-point GPA intervals and the 0 to 6 goal ratings. For grading systems it was based on the half-point GPA intervals and on the utility scores for grading systems recoded into intervals (0 to 36 = 1, 37 to 72 = 2, 73 to 108 = 3, 109 to 144 = 4, 145 to 180 = 5, 181 to 216 = 6, 217 to 252 = 7). Whenever the count of any row or column was less than 8, it was grouped with another row or column to prevent inaccuracies arising from low cell counts. For the same reason, GPA intervals below 2.00 were not included in any computations, since there were only two subjects in these intervals.

---

2Ibid., pp. 518-528.

3Ibid., pp. 347-351 and 529-532.

4Ibid., pp. 528-532.
CHAPTER IV

FINDINGS

The statistical analysis for all experimental hypotheses are presented below. Where results were significant, tables of the summary of analysis of variance and tables of the additional Newman Keuls analysis are presented and discussed for both sets of hypotheses. As an adjunct study, the experimental data was further analyzed according to student grade-point-average, using the Chi-Square technique. These results are also presented and discussed.

Hypotheses Concerning the Objectives of Grading

The results of the analysis of variance for grading objectives is presented below. There were no significant interaction effects.

Hypothesis One: "Grading objectives are most accepted by majors in professional areas and least accepted by underclassmen."

This hypothesis was not confirmed. Contrary to the hypothesis, underclassmen tended to respond more favorably to grading objectives than either upperclassman group, but the differences were not significant at $P < .05$ ($F=2.22; df=2.297; P > .05$). The mean rating of objectives for underclassmen was 3.35; that of upperclass arts and science students was 3.06; upperclass professional students was 3.02. A profile of these means and the analysis of variance are presented in Figure 1 and Table 1 respectively.

Hypothesis Two: "Certain objectives, such as the selection functions and mark of recognition for achievement, will receive greater acceptance than other objectives, such as
FIGURE 1

Mean Group Rating of Grading Objectives
TABLE 1

Summary of Analysis of Variance of Rating of Importance of Grading Objectives by Underclassmen, Upperclassmen--Arts and Sciences, Upperclassmen--Professional

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>P Less Than</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groups (A)</td>
<td>2</td>
<td>22.872</td>
<td>2.22</td>
<td>.25</td>
</tr>
<tr>
<td>Subjects Within</td>
<td>297</td>
<td>10.291</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objectives (B)</td>
<td>6</td>
<td>85.397</td>
<td>42.76</td>
<td>.01</td>
</tr>
<tr>
<td>AB</td>
<td>12</td>
<td>2.355</td>
<td>1.18</td>
<td>.30</td>
</tr>
<tr>
<td>B x Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Groups</td>
<td>1782</td>
<td>1.997</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
means of rewarding desired behavior and providing motivation." Significant differences at \( P < .05 \) between the mean ratings of the desirability of grading objectives were found, though not necessarily in the predicted order (\( F=42.76; \text{df}=6.1782; P < .01 \)).

The three objectives which were intrinsically related to the education process were significantly higher than all other objectives, but they did not differ among themselves. These were "feedback for instructor" (with a mean rating of 3.71), "feedback for student" (3.65), and "motivation for student" (3.56). "Selection for school" (3.09) and "mark of recognition" (3.06) were significantly lower than the first three objectives, significantly higher than the remaining two, and did not differ between themselves. "Reward for desired behavior" (2.33) and "selection for employer" (2.61) were rated significantly lower than all other objectives and did not differ between themselves. The analysis of variance and Newman-Keuls analysis are presented in Table 1 and Table 2 respectively. The profile of these means is presented in Figure 2.

Hypothesis Three: "Certain objectives will be more acceptable to one group of students than to the other, specifically: (a) underclassmen will consider motivation, and marks of recognition, particularly undesirable as grading objectives; (b) students in professional areas will consider selection for graduate school and employer, and means of rewarding desired behavior, particularly desirable as grading objectives." This hypothesis assumed that underclassmen and professional students will rate various grading objectives in opposite directions, while arts and science students will rank objectives somewhere in between the first two groups. No significant interaction differences were obtained, however, at \( P < .05 \), and group cell means did not follow the predicted relationships. The analysis of variance table is given in Table 1. The table of individual cell means is presented in Table 19 of Appendix B. The profile of these means is presented in Figure 2.

**Hypotheses Concerning Grading Systems**

An analysis of variance for simple main effects was conducted in the case of grading systems, since the interaction effect for the 3x6 (student groups by grading systems)
Table 2
Newman-Keuls Analysis of Overall Main Effect of Rating of Grading Objectives

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Differences Between Means</th>
<th>r</th>
<th>$9.95(r, 120) \sqrt{\frac{MS_{\text{error}}}{n}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>Mean</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>B4</td>
<td>2.33</td>
<td>--</td>
<td>0.28</td>
</tr>
<tr>
<td>B7</td>
<td>2.61</td>
<td>--</td>
<td>0.45&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>B5</td>
<td>3.06</td>
<td>--</td>
<td>0.03</td>
</tr>
<tr>
<td>B6</td>
<td>3.09</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>B3</td>
<td>3.56</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>B2</td>
<td>3.65</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>B1</td>
<td>3.71</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

P < .05 4 7 5 6 3 2 1 Objectives not differing significantly at P < .05 from each other share a common underline.

Objective B1 - feedback for instructor  
B2 - feedback for student  
B3 - motivation for student  
B4 - reward for desired behavior  
B5 - mark of recognition  
B6 - selection for school  
B7 - selection for employer
Mean Ratings of Grading Objectives by Three Student Groups
Hypothesis One: "More traditional grading systems will receive the greater acceptance." In the case of overall main effects ($F=67.44; \text{df}=5.1485; P < .01$) and simple main effects (Underclassmen ($F=32.69; \text{df}=5.1485; P < .001$), Arts and Science ($F=25.72; \text{df}=5.1485; P < .001$), Professional ($F=1439; \text{df}=5.1485; P < .001$)) for each group of students, significant differences were obtained between grading systems. A profile of means and the $F$-table for the analysis of variance of simple main effects are presented in Figure 3 and Table 4 respectively. A table of all means is presented in Table 20 of Appendix B. The more traditional grading systems were A-F (4 point) and A-F with P-F option. Least traditional were total P-F and descriptive grading. The grand mean for all grading systems was 73.47.

For each group the "total P-F system" was rated significantly lower than all other grading systems, as was predicted: (Underclassmen - 47.87, Arts and Science - 46.84, Professional - 49.78). On the other hand, "descriptive grading" had the highest mean rating for each group (Underclassmen - 92.70, Arts and Science - 90.09, Professional - 82.22). It was significantly higher than all other grading systems for arts and science students, and also significantly higher than all other grading systems except "A-F with + and -" for professional students. This outcome is contrary to predicted results. The traditional grading systems tended to be rated in the middle area, significantly higher than some grading systems and significantly lower than others: (A-F (4 point) - 74.75, A-F with P-F - 73.98).

The Newman-Keuls analysis of mean ratings of grading systems by the three student groups is presented in Tables 5, 6, and 7.
### TABLE 3

Summary of Analysis of Variance of Utility Values of Grading Systems by Underclassmen, Upperclassmen--Arts and Sciences, Upperclassmen--Professional

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>P</th>
<th>Less Than</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Groups (A)</td>
<td>2</td>
<td>22,540.04</td>
<td>2.61</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>Subjects Within Groups</td>
<td>297</td>
<td>8,642.72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grading Systems (C)</td>
<td>5</td>
<td>55,875.45</td>
<td>67.44</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>10</td>
<td>2,221.11</td>
<td>2.68</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>C x Subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Groups</td>
<td>1485</td>
<td>828.49</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE 3

Mean Ratings of Grading Systems by Three Student Groups
### TABLE 4

Summary of Analysis of Variance for Simple Main Effects of Grading Systems on Utility Scores of Grading Systems

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F</th>
<th>P Less Than</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grading Systems (C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>at Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C at A₁</td>
<td>5</td>
<td>27,085.88</td>
<td>32.69</td>
<td>.001</td>
</tr>
<tr>
<td>C at A₂</td>
<td>5</td>
<td>21,310.90</td>
<td>25.72</td>
<td>.001</td>
</tr>
<tr>
<td>C at A₃</td>
<td>5</td>
<td>11,920.73</td>
<td>14.39</td>
<td>.001</td>
</tr>
<tr>
<td>Within Cells</td>
<td>1485</td>
<td>828.49</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*A₁* - Underclassmen  
*A₂* - Arts and Sciences  
*A₃* - Professional
TABLE 5

Newman-Keuls Analysis of Simple Main Effect of Rating of Grading Systems by Underclassmen

<table>
<thead>
<tr>
<th>Grading Systems</th>
<th>Difference Between Means</th>
<th>r</th>
<th>$9.95(r, 120) \sqrt{\frac{MSError}{n}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>Mean</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>C5</td>
<td>47.87</td>
<td>--</td>
<td>33.94&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>C2</td>
<td>81.81</td>
<td>--</td>
<td>3.21</td>
</tr>
<tr>
<td>C1</td>
<td>85.01</td>
<td>--</td>
<td>0.23</td>
</tr>
<tr>
<td>C6</td>
<td>85.25</td>
<td>--</td>
<td>5.05</td>
</tr>
<tr>
<td>C4</td>
<td>90.30</td>
<td>--</td>
<td>2.40</td>
</tr>
<tr>
<td>C3</td>
<td>92.70</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>$p < .05$  

Grading systems not differing significantly at $P < .05$ from each other share a common underline.

$MS_{error} = 828.49$  
$df_{error} = 1485$  

Grading System:  
C1 - A-F with P-F  
C2 - A-C without D or F  
C3 - Descriptive grading  
C4 - A-F with + and -  
C5 - Total P-F  
C6 - A-F (4 point)
TABLE 6
Newman-Keuls Analysis of Simple Main Effect
of Rating of Grading Systems by
Arts and Science Students

<table>
<thead>
<tr>
<th>Grading Systems</th>
<th>Difference Between Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>Mean</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>C5</td>
<td>46.84</td>
</tr>
<tr>
<td>C1</td>
<td>65.77</td>
</tr>
<tr>
<td>C6</td>
<td>68.34</td>
</tr>
<tr>
<td>C2</td>
<td>73.71</td>
</tr>
<tr>
<td>C4</td>
<td>79.72</td>
</tr>
<tr>
<td>C3</td>
<td>90.09</td>
</tr>
</tbody>
</table>

<sup>a</sup><sup>p< .05</sup>  

Grading systems not differing significantly at \(P<.05\) from each other share a common underline.

\[
\text{MS}_{\text{error}} = 828.49 \\
\text{d}_{\text{error}} = 1485
\]
TABLE 7
Newman-Keuls Analysis of Simple Main Effect of Rating of Grading Systems by Professional Students

<table>
<thead>
<tr>
<th>Grading Systems</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>r</th>
<th>( 9.95(r, 120) \sqrt{\frac{MS_{error}}{n}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>Mean</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>C5</td>
<td>49.78</td>
<td>--</td>
<td>16.43</td>
<td>--</td>
<td>20.87</td>
<td>21.39</td>
<td>5</td>
</tr>
<tr>
<td>C2</td>
<td>66.21</td>
<td>--</td>
<td>--</td>
<td>4.44</td>
<td>4.96</td>
<td>8.87</td>
<td>4</td>
</tr>
<tr>
<td>C6</td>
<td>70.65</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>0.45</td>
<td>4.43</td>
<td>16.01</td>
</tr>
<tr>
<td>C1</td>
<td>71.14</td>
<td>--</td>
<td>--</td>
<td>3.91</td>
<td>--</td>
<td>11.05</td>
<td>4</td>
</tr>
<tr>
<td>C4</td>
<td>75.08</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>7.14</td>
<td>--</td>
<td>3</td>
</tr>
<tr>
<td>C3</td>
<td>82.22</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>2</td>
</tr>
</tbody>
</table>

\( a_p < .05 \)  
5 2 6 1 4 3  Grading systems not differing significantly at \( P < .05 \) from each other share a common underline.

MS_{error} = 828.49  
\( df_{error} = 1485 \)  

Grading System:  
C - A-F with P-F  
C - A-C without D or F  
C - Descriptive Grading  
C - A-F with + and -  
C - Total P-F  
C - A-F (4 point)
Hypothesis Two: "Grading systems will be more accepted by students in professional fields and less accepted by underclassmen." The overall main effects, while approaching significance, were not truly significant at $P < .05$ ($F=2.61; df=2.297; .05 < P < .10$). The results are given in Table 3. However, in two of six cases--"A-F with P-F option" ($F=4.63; df=2.1782; P < .01$) and "A-F (4 point)" ($F=3.94; df=2.1782; P < .02$)--the simple main effects indicated significant differences between groups. In two other cases--"A-C without D or F" ($F=2.86; df=2.1782; P < .06$) and "A-F with + and -" ($F=2.86; df=2.1782; P < .06$) near significance, less than the required value of $P < .05$ was obtained. Profiles of means and the $F$ table for the analysis of simple main effects are presented in Figure 4 and Table 8. In each case where significant differences were obtained, mean ratings by underclassmen were significantly higher than those by the other two groups, which did not differ from each other (A-F with P-F: Underclassmen - 85.01, Arts and Science - 65.77, Professional - 71.14; A-F (4 point): Underclassmen - 85.25, Arts and Science - 68.34, Professional - 70.65). Where results approached significance, the position of group means suggested a similar relationship (A-F with + and -: Underclassmen - 90.30, Arts and Science - 79.72, Professional - 75.08; A-C without D or F: Underclassmen - 81.81, Arts and Science 73.71, Professional - 66.21). Only in one instance (total P-F) did underclassmen fail to have the highest mean (Underclassmen - 47.87, Arts and Science - 46.84, Professional - 49.78), and the difference was insignificant. In general, it was concluded that underclassmen rate grading systems more favorably than either group of upperclassmen, in contradiction to the predicted result. The hypothesis was, therefore, not supported. The Newman-Keuls analysis of overall simple main effects of means of group ratings of grading systems by grading systems is presented in Tables 9 and 10.

Hypothesis Three: "More traditional grading systems will be most accepted by students in professional areas, and that less traditional grading systems will be most accepted by underclassmen." This hypothesis was not confirmed. Both underclassmen and professional students tended to rank grading systems in a similar order. There were no grading systems ranked significantly higher by one group but significantly lower by another group. The more traditional grading systems were, in fact, ranked significantly higher by
Mean Ratings of Grading Systems by Three Student Groups by Six Grading Systems
<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Degrees of Freedom</th>
<th>Mean Square</th>
<th>F Ratio</th>
<th>P Less Than</th>
</tr>
</thead>
<tbody>
<tr>
<td>Groups (A) At Grading System</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A at C₁</td>
<td>2</td>
<td>9867.38</td>
<td>4.63</td>
<td>.01</td>
</tr>
<tr>
<td>A at C₂</td>
<td>2</td>
<td>6086.83</td>
<td>2.86</td>
<td>.06</td>
</tr>
<tr>
<td>A at C₃</td>
<td>2</td>
<td>2976.23</td>
<td>1.40</td>
<td>.25</td>
</tr>
<tr>
<td>A at C₄</td>
<td>2</td>
<td>6085.09</td>
<td>2.86</td>
<td>.06</td>
</tr>
<tr>
<td>A at C₅</td>
<td>2</td>
<td>222.59</td>
<td>0.10</td>
<td>.90</td>
</tr>
<tr>
<td>A at C₆</td>
<td>2</td>
<td>8407.39</td>
<td>3.94</td>
<td>.02</td>
</tr>
<tr>
<td>Within Cells</td>
<td>1782</td>
<td>2131.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C₁ - A-F with P-F  
C₂ - A-C without D or F  
C₃ - Descriptive Grading  
C₄ - A-F with + and -  
C₅ - Total P-F  
C₆ - A-F (4 point)
TABLE 9

Newman-Keuls Analysis of Means of Groups Ratings
Grading systems for A-F With P-F

<table>
<thead>
<tr>
<th>Group</th>
<th>Difference Between Means</th>
<th>r</th>
<th>$9.95(r, 120) \sqrt{\frac{MS_{error}}{n}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Code</td>
<td>Mean</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>A2</td>
<td>65.77</td>
<td>--</td>
<td>5.37</td>
</tr>
<tr>
<td>A3</td>
<td>71.14</td>
<td>--</td>
<td>13.87$^a$</td>
</tr>
<tr>
<td>A1</td>
<td>85.01</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

$^aP < .05$

Grading systems not differing significantly at $P < .05$ from each other share a common underline.

$MS_{error} = 2131.56$ Group $A_1$ - Underclassmen
$df_{error} = 1782$ $A_2$ - Arts and Science
$A_3$ - Professional
TABLE 10

Newman-Keuls Analysis of Means of Group Ratings of Grading Systems for A-F (4 point)

<table>
<thead>
<tr>
<th>Group</th>
<th>Code</th>
<th>Mean</th>
<th>Difference Between Means</th>
<th>r</th>
<th>$9.95(r, 120)\sqrt{\frac{MS_{error}}{n}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A₂</td>
<td>68.34</td>
<td>-</td>
<td>2.31</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>A₃</td>
<td>70.65</td>
<td>-</td>
<td>14.60&lt;sup&gt;a&lt;/sup&gt;</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>A₁</td>
<td>85.25</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

P < .05 2 3 1 Grading systems not differing significantly at P < .05 from each other share a common underline.

$MS_{error} = 2131.56$  
$df_{error} = 1782$

Group A₁ - Underclassmen  
A₂ - Arts and Sciences  
A₃ - Professional
Underclassmen than by the other groups (A-F with P-F: Underclassmen - 85.01, Arts and Science - 65.77, Professional - 71.14; A-F (4 point): Underclassmen - 85.25, Arts and Science - 68.34, Professional - 70.65). No significant differences in mean ratings resulted for the non-traditional systems of descriptive grading or total P-F grading, nor was even a trend evident. The significant interaction between groups and grading systems resulted primarily from groups differing significantly from each other in regard to some grading systems and not others, and did not result from reversals in order of rankings.

Analysis of Ratings by Grade Point Average

As an adjunct study the relationships of students' grade-point-averages to their ratings of grading objectives and grading systems were explored. It was hypothesized that students would evaluate grading objectives and grading systems differently according to their varying success in achieving under the traditional system. The means of student ratings of grading objectives and grading systems according to grade point average intervals are presented in Tables 21 and 22 of Appendix B.

Chi square analyses were run to study the relationship for each objective and each grading system. The chi square table for grading objectives was based on half-point GPA intervals and the 0 to 6 goal ratings. For grading systems it was based upon half-point intervals, and the utility scores for grading systems recoded into intervals (0 to 36 = 1, 37 to 72 = 2, 73 to 108 = 3, 109 to 144 = 4, 145 to 180 = 5, 181 to 216 = 6, 217 to 252 = 7). Whenever the count of any row or column was less than 8, it was grouped with another row or column to prevent inaccuracies from low cell counts. For the same reason GPA intervals below 2.00 were not included in any computations, since there were only two subjects in these intervals.

In the case of grading objectives, seven chi square analyses were run. Significant relationships between students' grade-point-average and their mean rating of objectives was not found in any case. It is concluded that there is no relationship between the two. The results of the chi square analyses are presented in Table 11.
<table>
<thead>
<tr>
<th>Goals</th>
<th>Chi Square</th>
<th>df</th>
<th>P Less Than</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback for instructor</td>
<td>17.90</td>
<td>18</td>
<td>.50</td>
</tr>
<tr>
<td>Feedback for student</td>
<td>17.08</td>
<td>18</td>
<td>.70</td>
</tr>
<tr>
<td>Motivation for student</td>
<td>21.89</td>
<td>18</td>
<td>.30</td>
</tr>
<tr>
<td>Reward for desired behavior</td>
<td>8.45</td>
<td>18</td>
<td>.99</td>
</tr>
<tr>
<td>Mark of recognition</td>
<td>20.81</td>
<td>18</td>
<td>.30</td>
</tr>
<tr>
<td>Selection for school</td>
<td>20.98</td>
<td>18</td>
<td>.30</td>
</tr>
<tr>
<td>Selection for employer</td>
<td>19.74</td>
<td>18</td>
<td>.50</td>
</tr>
</tbody>
</table>
In the case of grading systems, six chi square analyses were run. Significant results were obtained in two cases, descriptive grading ($x^2=27.82, \text{df}=15, P < .05$) and A-F with + and - ($x^2=25.11, \text{df}=15, P < .05$). These two grading systems gave more precise or detailed information about academic achievement than any of the other grading systems. It is concluded that more able students with high grade-point-averages prefer more informational and accurate grading systems, and that no other relationships exist. The results of these chi square analyses are presented in Table 12. The complete chi square tables for both the descriptive grading and A-F with + and - analyses are presented in Table 13 and Table 14 respectively. In the individual cells of the tables are four figures which indicate the frequency of the cell, the percentage of frequency of the row represented by the cell, the percentage of the column, and the percentage of the total N, respectively.

**Summary of Findings**

Most hypotheses for this study were not supported by experimental results. There were, however, significant differences which resulted from the various analyses. Grading objectives which were intrinsically related to the educational process were rated higher than all others. These included feedback to instructor, feedback to student, motivation for the student. The objectives of rewarding desired behavior and serving as a selection tool for employer were ranked lower than all other objectives. Underclassmen rated grading systems higher than did either group of upperclassmen, who in their rating did not differ from each other. The total P-F system was least acceptable to students, while descriptive grading tended to be rated highest. Some additional differences in ratings of grading systems appeared within upperclassman student groups.

**Hypotheses concerning the objectives of grading:**

*Hypothesis One:* "Grading objectives are most accepted by majors in professional areas and least accepted by underclassmen." The hypothesis was not confirmed and the results, though not significant, were in the opposite direction.
### TABLE 12

Chi Square Analyses of Student Ratings of Grading Systems According to Grade Point Average Intervals

<table>
<thead>
<tr>
<th>Grading System</th>
<th>Chi Square</th>
<th>df</th>
<th>P Less Than</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-F with P-F</td>
<td>14.68</td>
<td>12</td>
<td>.30</td>
</tr>
<tr>
<td>A-C without D and F</td>
<td>6.15</td>
<td>12</td>
<td>.95</td>
</tr>
<tr>
<td>Descriptive Grading</td>
<td>27.82</td>
<td>12</td>
<td>.05</td>
</tr>
<tr>
<td>A-F with + and -</td>
<td>25.11</td>
<td>12</td>
<td>.05</td>
</tr>
<tr>
<td>Total P-F</td>
<td>6.74</td>
<td>12</td>
<td>.70</td>
</tr>
<tr>
<td>A-F (4 point)</td>
<td>16.30</td>
<td>12</td>
<td>.20</td>
</tr>
</tbody>
</table>
## TABLE 13

Chi Square Table for Descriptive Grading System

<table>
<thead>
<tr>
<th>Na</th>
<th>Grade Point Average</th>
<th>Row Total % of Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.00-</td>
<td>2.50-</td>
</tr>
<tr>
<td>Row %b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Column %c</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total %d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utility Score</td>
<td>12a</td>
<td>12</td>
</tr>
<tr>
<td>0-36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>29.3b</td>
<td>29.3</td>
<td>22.0</td>
</tr>
<tr>
<td>21.1c</td>
<td>13.5</td>
<td>13.0</td>
</tr>
<tr>
<td>4.8d</td>
<td>4.8</td>
<td>3.6</td>
</tr>
<tr>
<td>37-72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.6</td>
<td>41.1</td>
<td>28.6</td>
</tr>
<tr>
<td>28.1</td>
<td>25.8</td>
<td>23.2</td>
</tr>
<tr>
<td>6.4</td>
<td>9.2</td>
<td>6.4</td>
</tr>
<tr>
<td>73-108</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.1</td>
<td>38.6</td>
<td>32.9</td>
</tr>
<tr>
<td>21.1</td>
<td>30.3</td>
<td>33.3</td>
</tr>
<tr>
<td>4.8</td>
<td>10.8</td>
<td>9.2</td>
</tr>
<tr>
<td>109-144</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.0</td>
<td>38.3</td>
<td>23.4</td>
</tr>
<tr>
<td>14.0</td>
<td>20.2</td>
<td>15.9</td>
</tr>
<tr>
<td>3.2</td>
<td>7.2</td>
<td>4.4</td>
</tr>
<tr>
<td>145-180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33.3</td>
<td>22.2</td>
<td>18.5</td>
</tr>
<tr>
<td>15.8</td>
<td>6.7</td>
<td>7.2</td>
</tr>
<tr>
<td>3.6</td>
<td>2.4</td>
<td>2.0</td>
</tr>
<tr>
<td>181-216</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>37.5</td>
<td>62.5</td>
</tr>
<tr>
<td>0.0</td>
<td>3.4</td>
<td>7.2</td>
</tr>
<tr>
<td>0.0</td>
<td>1.2</td>
<td>2.0</td>
</tr>
<tr>
<td>Column Total</td>
<td>57</td>
<td>89</td>
</tr>
<tr>
<td>% of Total N</td>
<td>22.9</td>
<td>35.7</td>
</tr>
</tbody>
</table>

Chi Square = 27.82  df = 15  P < .05

a The frequency in each cell
b The % of the row total in each cell
c The % of the column total in each cell
d The % of the total N in each cell
### TABLE 14

<table>
<thead>
<tr>
<th>Utility Score</th>
<th>0-36</th>
<th>37-72</th>
<th>73-108</th>
<th>109-144</th>
<th>145-180</th>
<th>181-216</th>
<th>Column Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row %</td>
<td>20a</td>
<td>34.5b</td>
<td>20.7</td>
<td>12.5</td>
<td>29.3</td>
<td>11.1</td>
<td>57</td>
</tr>
<tr>
<td>Row %</td>
<td>17</td>
<td>29.3</td>
<td>33.7</td>
<td>13.5</td>
<td>29.3</td>
<td>13.3</td>
<td>89</td>
</tr>
<tr>
<td>Row %</td>
<td>15</td>
<td>25.9</td>
<td>17.4</td>
<td>13.0</td>
<td>21.1</td>
<td>5.8</td>
<td>69</td>
</tr>
<tr>
<td>Row %</td>
<td>6</td>
<td>10.3</td>
<td>11.8</td>
<td>23.5</td>
<td>26.3</td>
<td>11.1</td>
<td>34</td>
</tr>
<tr>
<td>% of Total N</td>
<td>58</td>
<td>23.3</td>
<td>25.7</td>
<td>16.5</td>
<td>7.6</td>
<td>3.6</td>
<td>249</td>
</tr>
</tbody>
</table>

**Chi Square Table for A-F With + and - Grading System**

<table>
<thead>
<tr>
<th>Grade Point Average</th>
<th>2.00-</th>
<th>2.50-</th>
<th>3.00-</th>
<th>3.50-</th>
<th>4.00</th>
<th>% of Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row Total</td>
<td>58</td>
<td>23.3</td>
<td>25.7</td>
<td>16.5</td>
<td>7.6</td>
<td>3.6</td>
</tr>
<tr>
<td>Column %</td>
<td>2.49</td>
<td>2.99</td>
<td>3.49</td>
<td>4.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \text{Chi Square} = 25.11 \quad \text{df} = 15 \quad P < 0.05 \]

* a The frequency in each cell
* b The % of the row total in each cell
* c The % of the column total in each cell
* d The % of the total N in each cell
Hypothesis Two: "Certain objectives, such as the selection functions and mark of recognition for achievement, will receive greater acceptance than other objectives, such as means of rewarding desired behavior and providing motivation." Significant differences were found, though not necessarily in the order predicted. The three objectives intrinsically related to the learning process, "feedback for instructor," "feedback for student," and "motivation for student," were rated highest. "Selection for school" and "mark of recognition" were rated in the middle. "Reward for desired behavior" and "selection for employer" were rated lowest.

Hypothesis Three: "Certain objectives will be more acceptable to one group of students than to the other, specifically: (a) underclassmen will consider the grading goals, motivation and mark of recognition, particularly undesirable; (b) students in professional areas will consider the objectives, selection for graduate school and employer and means of rewarding desired behavior, particularly desirable." This hypothesis was not confirmed, as no significant results were obtained.

Hypotheses concerning grading systems:

Hypothesis One: "More traditional grading systems will receive the greater acceptance." This hypothesis was only partially confirmed as some, but not all, traditional systems were rated high, and some, but not all, non-traditional systems were rated low. Rather, grading systems which provide the most information, "descriptive grading" and "A-F with + and -" were rated highest, and the one grading system providing the least information, "total P-F system," was rated lowest.

Hypothesis Two: "Grading systems will be most accepted by students in professional fields and least accepted by underclassmen." The hypothesis was not confirmed. Where significant differences were obtained, underclassmen rated grading systems higher than did upperclassmen.

Hypothesis Three: "More traditional grading systems will be most accepted by students in professional areas, and the less traditional grading systems will be most accepted by underclassmen." The hypothesis was not confirmed. Grading
systems tended to be ranked in the same order by all groups, and underclassmen ranked traditional grading systems (as they tended to rank all grading systems) higher than did the other groups.
CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Background for the Study

The purpose of the study was twofold. From the literature an attempt was made to determine: (1) actual grading practices; (2) proposed functions or decision-making objectives served by grades; (3) the effectiveness of grades in serving these functions. From student opinion the study attempted to determine: (1) the relative importance or desirability of the proposed functions or decision-making objectives supposed to be served by grades; (2) the overall measure of the utility of various grading systems in attaining these functions; (3) the differences in judgment among selected groups of students. The specific hypotheses are presented with the summary of findings and conclusions below.

The Study Methodology

Experimental procedures included development of a questionnaire to sample student opinion, which included questions on grading objectives and grading systems using a seven unit Likert-type scale. Student ratings of objectives, and of the effectiveness of various grading systems in achieving those objectives, were combined to form a measure of overall grading system utility.

A pilot test was conducted to test and refine the questionnaire. Data from the study was used to obtain an estimate of error variance, from which the number of subjects needed in the study proper was calculated.

Three hundred volunteer students—one hundred underclassmen from the University College, one hundred upperclassmen majoring in professional fields, and one hundred
upperclassmen majoring in the arts and sciences--participated in the study. The hypotheses were tested using a fixed-factor, two-factor analysis of variance with repeated measures on the second factor. Where significant interaction effects were obtained a further analysis of variance for simple main effects was performed. Significant results were further analyzed, using the Newman-Keuls Procedure. As an adjunct study, the relationships of students' grade-point-averages to their ratings of grading objectives and grading systems was explored using the Chi Square statistic.

Summary of Findings and Conclusions

Results of the tests of hypotheses and interpretations were:

Concerning the objectives of grading:

1. **Hypothesis One:** that grading objectives are most accepted by majors in professional areas and least accepted by underclassmen.

   Upperclassmen did not respond more favorably than underclassmen to grading objectives. The latter's response was somewhat more favorable, but the differences were not significant.

   It was expected that upperclassmen would generally react more favorably to the objectives of grades, having been socialized to a much greater degree into the system than the underclassmen, and therefore, would accept more readily the stated purposes of grading. That this was not the case may be due to several reasons. The grading system at the secondary level and perhaps earlier may be sufficiently similar to the college level so that all basic attitudes and values have already been established, and as a result little additional change might be expected at the college level. Second, socialization into the system may not mean acceptance of grading objectives. Students may understand the grading system and its implications for future study and employment and effectively work within this framework.
while still not accepting to any greater degree the basic assumptions of the system. It might further be theorized that their experience would make their judgment more critical. Whatever the combination of effects, student judgment about the validity of grading objectives remains relatively stable in the undergraduate years.

2. **Hypothesis Two:** that certain objectives such as the selection functions, and mark of recognition for achievement will receive more acceptance than other objectives such as means of rewarding desired behavior and providing motivation.

Some grading objectives differed significantly from others in acceptability. The differences were not necessarily in the predicted order. Reward for desired behavior and selection for employer were significantly lower than all other objectives.

The three functions most closely related to the teaching-learning process, feedback for instructor and for student, and motivation, rated significantly higher than all other functions. The intrinsic relation of grading objectives to the teaching-learning process appears to be the primary determinant of student judgment in this area.

It should also be noted that on an overall basis students did not simply take a negative attitude toward all objectives suggested as justification for grades. On the scale used in the study, three represented a neutral response. The overall mean was slightly on the positive side at 3.14. The three most favored goals were rated 3.71, 3.65, and 3.56 respectively.

3. **Hypothesis Three:** that certain objectives will be more acceptable to one group of students than the others, specifically: (a) underclassmen will consider motivation and mark of recognition particularly undesirable as grading objectives; (b) students in professional areas will consider selection for graduate school and employer and means of rewarding desired behavior particularly desirable as grading objectives.
No evidence was found to indicate that any interaction effect occurred between student groups and specific goals. These results indicate that by and large the student groups studied judge grading objectives in similar ways. It is still possible that a finer breakdown of student groups would result in differences, since it is reasonable to expect that art or education majors might judge differently from science majors, even though such differences fail to show up when students are integrated into larger groups.

Concerning grading systems:

1. **Hypothesis One:** that the more traditional grading systems will receive the greater acceptance.

   Significant differences were found in student judgment of the overall effectiveness of various grading systems in achieving the objectives of grading, as weighted by the perceived performance of the various functions. The hypotheses were confirmed in part. The results indicated, however, that those grading systems which provided most information, such as descriptive grading and A, B, C, D, F with + and -, were rated highest, while those providing least information, such as complete pass-fail, were rated lowest. Traditional as opposed to non-traditional grading systems produced no clear-cut results.

   It was expected that students who were socialized into the whole context of traditional grading would tend to favor those systems which were more traditional, but both descriptive grading, which is non-traditional, and the A-F with + and -, which is close to traditional, were highly favored. Total pass-fail, highly non-traditional, found little favor.

   In summary, the adequacy of information provided by a traditional or non-traditional system appears to be the primary factor in student judgment. The further analysis of this data by Chi
Square indicated that the more able student, as measured by grade-point-average, favored the more informational systems.

According to the scale used, a neutral rating for grading systems is 63. That is, if a student rated all objectives as well as the effectiveness of a grading system in achieving these objectives in the neutral area of the scale, then the resulting utility score would be 63. (See Chapter III, page 198, for derivation of rating scale). Again the overall judgment of students was positive, with the mean for all systems being 73.47. The two most favored systems were rated 88.34 and 81.70, while three other systems were given positive rating or approximately 74; only one was rated negatively.

2. Hypothesis Two: that grading systems will be more accepted by students in professional fields and less accepted by underclassmen.

The overall ANOVA did not indicate that any of the groups were more or less favorable in their judgment of grading systems. The ANOVA for the simple main effects indicated significant differences for A-F with the pass-fail option, and for A-F. In the case of A-C without D or F, and A-F with + and -, results approached but did not reach significance.

In general, the underclassmen gave grading systems more favorable ratings than those given by either group of upperclassmen, who did not differ among themselves. In each significant result underclassmen ratings were higher, and in only one case, where the groups ratings were almost identical, did the underclassmen fail to have the highest means. This trend is in the opposite direction of what was hypothesized. Possible reasons for these results are similar to those given for the first hypothesis relating to grading objectives.

3. Hypothesis Three: that more traditional grading systems will be most accepted by students in
professional areas and that less traditional grading systems will be most accepted by underclassmen.

No interaction effect was found between grading system and student groups. All groups tended to rank the grading systems in a similar order. As was the case for grading objectives, this study did not break down student response into fine groups. While student judgment does not differ on the whole in regard to grading systems, it should not be assumed that certain specific groups of students do not differ in their reactions to specific grading systems.

Concerning the ratings of students by grade point average:

1. There were no significant relationships between students' grade-point-average and their rating of grading objectives.

2. Students' grade-point-average was significantly related to their rating of grading systems in two instances: students with higher averages rated highest the two most informational grading systems, "descriptive grading," and "A-F with + and -.

**Summary of Grading Literature**

**Historical Overview**

The history of grading provides some considerable perspective upon present day concerns and at least to some extent ought to lead debate in productive directions by avoiding the same circles: many current topics of debate were already considered problems more than a hundred years ago.

The earliest evaluation of academic achievement in colonial colleges was based upon oral examinations on a common curriculum. Recording of the level of achievement was made by means of descriptive words or phrases and honorary tasks at commencement. By the middle 1800's, the written examination had come into wide usage, being considered more reliable because all students responded to the same questions
and the results could be reviewed by many. In the second half of the Nineteenth Century, numerical scales (usually based on 100) replaced the descriptive ratings, though at least one institution had for a time used a pass-fail system. A slightly more precise pass-with-distinction, pass, and fail system also in use was later modified by adding + and -.

By 1900 there was a strong shift away from the numerical scale of 100, which was considered more precise than could be justified, since it inaccurately implied some absolute percentage of knowledge. In consequence, the A, B, C, D, or F system was adopted on the general assumption that it provided the proper number of divisions by which students could be classified, a rationale largely negated when colleges began to compute numerical grade-point-averages to the hundredth of a unit from the letter grades.

By this time behavior or conduct was generally excluded as a factor determining grades. The merit of grades as rewards and as motivation was debated without resolution. The introduction of the elective system had an important result in that it led to examinations and the assignment of grades by the individual instructor on a course by course basis.

In the first third of the Twentieth Century, the "scientific" approach to grades was the fashion. Objective testing was widely adopted and the traditional letter grades were carefully defined according to the normal curve. The inaccuracies of grading were reported in the literature, with suggestions for correcting the deficiencies by modifying grade distributions. There was also debate and concern about various factors which motivated instructors in assigning grades, and especially about their lack of consistency. A few critics suggested abolishing grades entirely.

Grading Systems

A large number of grading systems are in use and have been proposed. The dominant system at present is the A, B, C, D, F system; which has been modified in various significant and insignificant ways. One very popular modification has been the adoption of a limited pass-fail option; another modification has been the attempt to increase precision by
the use of + and -. Innumerable other variations of the system exist, such as counting only the second grade when a course is repeated.

Grade distributions based upon a normal curve of some sort have been frequently discussed but are rarely strictly followed. Generally each instructor has had a favored distribution of grades. Methods for adjusting grade distributions for student ability have been formulated but rarely used formally or incorporated into a grading system. Normative grading based simply on student rank has been proposed, but has not been used as a formal grading system. Forms of contract grading have been used by some individual instructors in the context of a traditional grading system and involve the specification of achievements with students on an individual basis.

Some writers, arguing that academic achievement is complex, have suggested using multiple dimension grading, in which several facets of student performance are recorded in symbols, or descriptive grading, in which several paragraphs of prose may be used to describe various dimensions of student achievement. A small number of institutions have used descriptive grading. Some faculty have allowed students simply to grade themselves; understandably, the result is generally high grades. No institution has adopted this approach as policy.

Sometimes complete pass-fail systems have been modified to provide two levels of satisfactory work—honors or pass—which is only one step removed from traditional grading. Non-punitive grading has been growing in popularity in recent years and provides only for the recognition of satisfactory work. Several critics of grades have proposed abolishing them, but this has not happened anywhere. Individualized grading systems were also proposed whereby a single institution might maintain several of the grading systems described above and use the particular system which best answered an individual student's needs. The literature does not record the implementation of such a plan, probably as a result of threatened resignations of all registrars.

Finally, it is possible to combine facets of many grading systems such as pass-fail with description, but despite all the discussion and possibilities, the dominant form of grading remains the traditional A to F with a limited
pass-fail option. There are, of course, minor differences from institution to institution, but these do not affect the reality of grading in any significant way.

Objectives of Grading

Weighing the objectives or purposes of grades is essential to any consideration of grading, since they are ultimately the criteria by which the whole process does or does not make sense. Again, depending upon objectives, the grading process itself might be structured in various ways. Unfortunately the literature does not indicate general agreement about objectives. Therefore it follows that little agreement about desirable forms of grading, grading outcomes, and most other matters relating to grades can be reached, or that such discussion will be very productive. There is some general acceptance of certain intrinsic objectives, such as motivation for the student, feedback to the student and to the instructor on the student's progress, and guidance to the student in making future study plans, but even here some argue that these purposes can well be achieved without grades as such.

Other objectives less intrinsic to learning are hotly disputed. These include grades as providing for selection by graduate schools and prospective employers; as a basis for awarding credit and honors and for continuance in school; as a measure of the efficiency of instruction, and as a preparation for life in a competitive system.

Determinants of Grades

What factors instructors actually use in assigning grades to students is also a frequent area of disagreement. For example, determining grades in college by traditional measures of academic ability has been explored intensively. The literature indicates that while significant correlations in the .70's have been found for high school rank, this explains only about half of the causal factors associated with the college grade-point average. Purer ability measures, such as SAT or ACT, correlate in the .50's and explain only one fourth of the causal factors. Clearly many other dimensions which have not been well explained are involved.
Ability or talent should be considered as multi-dimensional, and the various dimensions may not be closely correlated. Depending upon what array of talents is required in a particular task, an individual might do well or poorly. The literature does show that high creativity tends to be negatively correlated with grade success, though it may be that in some areas this is not so. Interpersonal skills and student-faculty interaction affect grades in subtle ways not always obvious on a common sense basis. Student personality traits as well as sex also affect grades.

Obviously then, the specific criteria consciously considered by faculty vary considerably, for though achievement is always said to be primary, what constitutes achievement is not always easy to determine. The student's progress over a semester, his effort, attendance, behavior, achievement in the light of his potential, and his class participation all are admittedly factors at various times and in various ways. While it is not reasonable to expect a standard set of criteria to be used in the same way at all times in every field of study, considerable movement in this direction is most desirable.

Grading Outcomes

The outcomes of traditional grading are difficult to evaluate in many areas a) because not a great deal of attention in the literature has been focused on this area in a substantive and scientific way; and b) because of the difficulty in isolating the effects of what grading represents from many confounding variables.

The selection function for graduate and professional schools must be considered a fairly effective outcome of grading, long been used extensively for this purpose. Significant positive correlations have been found, in spite of constricted ranges of ability. Selection and screening for employers must be considered unsuccessful. Correlations, where found to exist, are so low that they might be very easily ignored. Grades which represent a multi-dimensional and complex reality are simply not satisfactory measures of the array of talents required for adult success.
Several functions, such as reporting on academic progress to students and parents, and serving as a basis for granting credit and various honors, are well handled by grades. This may or may not be considered a significant accomplishment, depending upon whether or not one admits that grades represent real and significant achievement.

Whether the teaching-learning process is facilitated by traditional grades is a matter of strong disagreement, with some evidence available which bears on this point. Grades are unquestionably a very strong motivating factor for students. Almost without exception, students strive for some level of satisfactory grades, this level being defined by each individual student for himself. In cases where grades were effectively eliminated in a course or where their impact was minimized, the level of student achievement was lowered.

The concern of the student for satisfactory grades pervades the teaching-learning process. Results can be dysfunctional, in creating excessive stress, which adversely affects performance, or in motivating students to activities which lead to a higher grade without higher achievement. These latter activities would include such things as cheating, taking "easy" courses, "psyching out" the instructor. Likewise, where an instructor's actual means or basis for assigning grades does not coincide with intended objectives, for example, when tests focus on details, students will neglect the significant objectives in favor of those activities rewarded by the instructor. There is strong opinion and some evidence to indicate that these dysfunctions occur to some extent. Nevertheless they are not intrinsic results of grading but rather the result of some fault in evaluation and in grading. The motivation resulting from grades generally must be considered a positive factor for the large majority of students.

The value of the feedback to students and instructors is questionable, and one's position has to follow the definition of grading. Narrowly defined as the assignment of a symbol at the end of a term to signify level of achievement, grading provides almost no feedback information to student or instructor which is both diagnostic and timely as it must be if student learning is to be facilitated. Defined more broadly as the entire range of activity upon which a final course grade is ultimately based, grading of
course can provide for the feedback functions very well, de-
pending upon the nature and timing of the testing and eval-
uation methods. It is also true that this feedback function
can be served well, even if the final results are never sum-
marized into some sort of course grade.

A point completely neglected in the literature is the
impetus given to the individual instructor to evaluate stu-
dent achievement carefully as a result of the necessity for
giving a final course grade which can be defended or logi-
cally explained to the student. Whether instructors would
in fact fulfill the feedback functions without this incen-
tive is an empirical and pragmatic matter which has not been
explored.

It is argued that competition inherent in grading is
useful in preparing students for a competitive world but
there is no empirical evidence to demonstrate this claim,
nor is there sufficient rationale for using the competitive
world (which in many cases is not competitive) for a model.
More important, perhaps, is the fact that there is also some
concern lest excessive competition result in educational
dysfunctions.

The grading system determines the social context for
students and faculty in important ways: the student's self-
image and sense of success, for example, are based upon the
ability to secure satisfactory grades. Furthermore, the
grade-point-average affects the student's relations with
peers and the "fairness" of an instructor in assigning
grades and the level of the grades assigned affect student-
faculty interaction. Not only is student activity to a
large extent determined by grades, faculty attitudes and
behavior likewise relate to some extent to grades assigned.
The upholding of standards and of disciplinary ideals can
be traced to demanding instructors who assign few high
grades though other instructors, wishing to encourage and
support students, assign many high grades. Thus different
grading philosophies can be associated with different dis-
ciplines.

Student attitudes toward grading have been mixed.
In the late '60's, a sizable minority of students was re-
ported as desiring the abolition of grades but this minor-
ity has decreased in recent years. In some institutions
student opinion has been favorable to selected innovations,
and in general is strongly in favor of the limited pass-fail option, which has been widely adopted. That various dysfunctions such as cheating, and studying for tests rather than for learning, do result from grading is generally agreed upon by student opinion.

Many of these same dysfunctions of grades are repeatedly mentioned in the literature, though their seriousness or impact has not been established in any empirical way. One such dysfunction, the displacing of ends (learning) by means (grades), is frequently noted. To the extent that tests and other measurements do not coincide with desired learning, or to the extent that activities not related to learning, i.e. cheating, are encouraged, this point has merit.

The extrinsic motivation provided by grades is viewed by some as undesirable. Yet while it is probably true that learning for the love of learning is most desirable, it does not necessarily follow that grading eliminates such motivation, nor is it clear that extrinsic motivation is necessarily bad, since it does seem to raise student achievement levels. Even faculty who teach for the love of teaching expect certain extrinsic rewards.

A more serious criticism of traditional grading relates to the meaning of grades. It is obvious that the scope of information provided by a letter symbol is limited, and even that limited information depends to a large extent upon the idiosyncrasies of the individual instructor, whose grading may be determined by a large variety of factors used in different combinations, with various emphases or weights. Obviously there is a strong need for grades to be carefully defined, both on a general level for an entire institution and specifically for each department or field of study.

A further variety of dysfunctions has been suggested in the literature, but again validity of these concerns has not been empirically established. Grades, for example, may or may not cause a role conflict for an instructor who is expected both to assist the student in learning and also to judge, and may discover the roles are contradictory. This problem, some research has found, could be resolved by the use of external examiners. Other dysfunctions noted in literature include the encouragement of a compartmental mentality, the possibility of inducing a negative self-image, and depersonalization.
The limited pass-fail option, which has received as much or more attention in the literature than any grading topic, has also had the benefits of some empirical study. In general, the objective of the option, to encourage students to explore new areas of study, has not been accomplished, and the level of student achievement in such courses is lower than elsewhere. The option, however, does reduce pressure on students (motivation according to critics of pass-fail), and allows them to have some freedom in the course instead of merely following the directions of the instructor.

Systems and Evaluation

It seems apparent that if any general improvement in the system of grading is to be effected a theoretical context and a basis for consideration of changes are required. Systems and evaluation theories provide such a context. The systems theory attempts to consider reality in all its complexity, studying components, structure, processes, inputs, outputs and feedback, the larger system, and objectives or functions. Grading is a subsystem of the curriculum, which is a subsystem of higher education, which is a subsystem of society, and society provides various resource inputs into higher education and expects, as one primary output, graduates who possess certain occupational as well as liberal skills. Society also expects that these graduates will have some sort of certification as to their accomplishments. It is unlikely that higher education in general is free to deviate from such constraints, though some individual institutions may do so.

In order to assist students to achieve objectives based on the expectations of society, departments have been established according to areas of study and curricula formulated by course units.

According to systems theory and the methodology of instructional systems, the overall objectives of a particular curriculum and its subdivisions must be very carefully defined. If one follows the CIPP evaluation model, it is then necessary to analyze inputs which are, for example, students entering a program with their varying abilities, interests, and prior study. The instructor is then able to formulate the components and structure of the learning
process and, through process evaluation, he is able to restructure the learning process, if satisfactory progress toward desired objectives is not being made. At the end of the learning sequence or course, he undertakes a final evaluation of student achievement, using the originally specified objectives as a standard of comparison. Such process evaluation provides feedback information to the student as well as the instructor, so that both can make adjustments, but the final, summative, or product evaluation, coming at the end of the learning process, cannot assist learning. Grading at this final point consists of symbolizing the individual student achievement on an absolute basis, by reference to the original objectives as standards, or by comparison with fellow students, perhaps without reference to actual or absolute levels of achievement. While process and product evaluation are in a sense distinct from grading, product evaluation is a prerequisite for grading, and it can be strongly argued that grading motivates the instructor to do a more complete and adequate job of evaluation, which is essential for effective learning.

Normative and Criterion Reference

In recent years considerable debate has occurred between the proponents of criterion and of norm referenced grading systems. According to evaluation theory, some comparison to a standard is necessary. Criterion reference suggests judgment of student achievement should be in comparison with carefully specified curricular objectives; normative reference implies comparison to the achievement of other students. Ultimately some accommodation is possible between the two viewpoints. Those using criterion reference must in the beginning adopt some basis for determining comprehension and level of depth, and the difficulty of the specified goals to be used as the standard for comparison. This is no doubt done on the basis of the instructor's experience with relative levels of achievement desired in specific areas.

Institutions which use normative reference maintain that the average achievement of large numbers of students tends to be constant, and that therefore using that average as a standard of comparison for grades will provide a more reliable indicator of actual student achievement.
While the long-run goals of both approaches are the same, normative reference incurs specific problems in the short-run. Under that system, for example, students in a small group are in effect playing a zero-sum game, in which the success of another student is prejudicial to their own success. Competition rather than cooperation is stressed, and grading literature suggests that the outcomes of such a system can be dysfunctional in a variety of ways. These problems are minimized to the extent that normative grades do correlate with absolute achievement; but to that same extent criterion and normative referenced grading converge. In addition a very important further advantage of criterion referenced grading is the emphasis placed upon the specification of desired outcomes. When this has been adequately done and evaluation is based upon these carefully specified outcomes, then the student knows clearly what is desired, and the less concrete and higher level objectives of a course are less likely to be overlooked. This approach also facilitates such approaches as teaching or learning for mastery, which attempt to minimize some of the traditional constraints upon the learning situation.

Technical Issues

One of the most important concerns and criticisms of traditional grading has been the inconsistency of standards used in the assignment of grades. The literature indicates beyond any doubt that this complaint is valid, for research proves that the relative number of high and low grades varies greatly no matter what unit of comparison is used, be it based on institutions, divisions, departments, level of study, or instructor. Further, these differences cannot be said to be a result of differing student ability levels, since studies controlling for this factor suggest even greater variability as often as not.

Established grading patterns are not variable, however. On a year-to-year basis instructors and departments who tend to grade high continue to do so, and vice versa, grades being part of personal, institutional, and social value systems which change only slowly. It has been established that overall average grades or grading patterns in the last few decades have been increasing significantly, and a number of reasons relating to values have been suggested.
Precision in grading is not a well-understood matter, judging from the literature. Critics concerned about the admitted inaccuracy of grading, not realizing that grosser categories create greater inaccuracy, have suggested reduction to two categories of pass and fail. With increasing disuse of D and F, traditional grading has basically become a three level system. In order to avoid significant rounding errors, at least ten categories are desirable. Further, there is some evidence to indicate that the gross categories presently in use tend to promote a statistical gaming approach to learning by students aiming to achieve better grades, though they may mean no overall increase in achievement.

The reliability of traditional grades has been shown to be moderate—which is perhaps somewhat more than expected in the light of the great inconsistency in factors determining grades, in the use of these factors in determining grades, and in grading standards. Correlations derived in various ways range generally in the .50 to .70 area and may explain about 40 percent of the true variance. It is clear that if appropriate precision were secured, along with the clarification of criteria, standards, evaluation and measurement procedures, the reliability of grades would be considerably enhanced.

Discussion and Recommendations

The review of the literature and results of the study have a number of serious implications in regard to present grading practices and future directions.

The Unsatisfactory Present

It should be clear to anyone familiar with the literature of grading that a number of serious problems presently beset grading and that they are ill-served by benign neglect. Among the many dysfunctions which have attracted the attention of the literature of grading are the distortion of ends and means, the narrowly-limited scope of information provided by grades, negative motivation, the induction of a compartmental mentality, the generation of negative self-image among students, role conflict for faculty, the inducement for students to study according to the
idiosyncracies of the instructor, the neglect of higher and more important goals of education for trivia, and many others. While many of these concerns are not "scientifically" established, the frequency of expression of concern by thoughtful and competent educators regarding them in itself should be sufficient reason for considering grading reform.

It is also clear from the literature that the concern students feel about getting grades and faculty in giving them are a central reality in higher education and exercise a pervasive influence upon the teaching-learning process. It appears obvious that grading has become a central ingredient in the attitude-value complex that governs both student behavior and faculty teaching styles and methods. Consequently, the manner in which grades are assigned has greater impact on the achievement of educational objectives than any other educational innovation or development.

Having noted the central role of grades and their admitted dysfunctions, it also must be added that great disagreement exists about the very objectives of grading. Neither individual institutions nor larger educational associations have attempted to formally define such objectives, much less to enforce any definition. In practice, instructors tend to follow their own self-determined purposes. Perhaps as a result, or at least as a correlate of undefined objectives, there have been no formal definitions of what a grade should represent, and so instructors use a variety of factors in inconsistent ways in determining grades. But not only is the system plagued with chaotic grading criteria; in addition, the standards and grading distributions of instructors, departments, institutions (or whatever the unit of comparison) vary significantly, and in many cases drastically.

Grades do enjoy modest reliability (though less than well-prepared objective evaluations); they have significant motivational effects upon students to learn more, and can be used fairly well in predicting success in future study. They are not effective, however, in predicting adult achievement, and they may or may not be useful in a feedback function to the student and instructor. On the whole they are used as a form of currency within an institution in the granting of credit, degrees, honors, etc., a currency demanded by the external systems with which higher education interacts.
In consideration of the ill-defined objectives of grading, of chaotic grading criteria, and of the serious matter of such dysfunctions produced by grading as varying standards, the critical role grades play in teaching and in learning, and the legitimate or illegitimate uses made of grades, a major and systematic attempt at grading reform is essential.

Required Scope and Level of Grading Reform

Now a critical gap presently exists in the initiation of any comprehensive approach to grading reform, the lack of knowledge about grades. If all opinion articles, survey articles, pass-fail grading articles, and correlation articles between high school, college, and graduate work were eliminated, little grading literature would remain. That residual literature and the meager experimental findings would hardly form a basis for reform, since their scope is limited and these studies--so many based on only a few classes or students--are highly susceptible to the Hawthorne effect.

A major concern of foundations interested in higher education as well as of other agents of funding, such as the state and federal government, should be to initiate and stimulate sound and critical experimental studies of grading on an institution-wide and interinstitutional basis. Through the introduction of alternative grading systems and evaluation techniques on a regular basis in an entire institution over a period of several years, and by the use of externally prepared achievement measures to study student accomplishment at both experimental and control institutions, some useful empirical evidence might be obtained for the first time, information which would provide a rational basis for grading decisions. To date even those few institutions which have introduced comprehensive new grading and evaluation systems, can only estimate their results by student and faculty opinion questionnaires. Large scale empirical studies are necessary before substantive progress can be made.

Another reason that makes grading a problem of national scope is the fact that an individual institution simply cannot change its own grading practices because of the restrictions arising from interconnectedness with other institutions in higher education and with the larger system
outside. For example, an institution which did not inflate its grading practices as most institutions recently have done, would find its graduates judged less competent, relative to graduates of other institutions, in seeking admission to graduate school and employment opportunities. Likewise, a school initiating any significant non-traditional grading system would have similar problems of articulation. Since the problems listed above are typical of all colleges and universities rather than just of a minority, grading reform definitely is a national concern and deserves to be considered on a large-scale basis.

It is recommended that a commission on grading and evaluation in higher education be established to focus on grading reform and to encourage systematic experimental study of grading. In due time, such a commission, as a result of its study, should formulate a policy position on grading which would include a statement defining the appropriate objectives of grading, the proper uses of grades, the criteria of a good grading system, and the role of evaluation as a part of any grading system.

It is then recommended that the ensuing policy be enforced on a national basis by means of accrediting associations, with each institution required to formulate a formal policy for evaluation and grading and to indicate how that policy is being implemented, as well as evaluating or stating its rationale in the context of the national policy. Thus an institution would be free to develop a grading policy in accord with its own needs providing it could be justified, yet under this approach any grading system would necessarily be a carefully developed system with appropriate rationale, and departures from usual systems, perforce, would be even more carefully examined. Each college should be required to describe its grading system accurately in all college literature and with each transcript issued; where it would be appropriate, depending upon the nature of the grading system, each college should also send with each transcript data on grading distributions at the college and list what uses of grades are legitimate.

Elements of Future Grading Reform

Although it is not possible to make many definitive statements about an ideal grading policy, the literature and
various studies do suggest strongly some directions which are described below.

**Systems and Evaluation.** It is highly recommended that systems and evaluation theory be used to provide a suitable framework for the development of better grading systems and practice. This approach is absolutely necessary in developing a national policy on grading, since the relations between institutions and between higher education and society, and the internal operations of institutions and the teaching-learning process, all must be considered.

System theory, with the conceptual framework of purpose and objective, structural elements and organization, inputs and outputs, and feedback systems is adequate for this task. Evaluation theory in the context of system theory is a suitable basis for the application to teaching and learning. The distinction between process and product evaluation is critical. The importance of the objective, the subsequent understanding of the comparison of outcome to desired objective, and the relativity or absoluteness of the comparison provide an extremely important framework for instructional methodology as well as grading.

Many of the endless arguments about grading spring from an unclear understanding of these matters. Product evaluation can be conducted separately from and without any process evaluation. On the other hand, process evaluation in higher education more often than not becomes the basis for most product evaluation, and this is the cause, at least in part, of many grading dysfunctions. The absence of carefully worked out objectives, essential in systems and evaluation theory and typically absent from many courses, is responsible for many other dysfunctions. Such a framework of objectives would assist students to understand and accept the legitimate objectives of grading, which objective may only partly coincide with those held by students and reported in the study.

**Normative and Criterion Reference.** Student performance ought not to be judged on a relative basis. Carried to its theoretical conclusion and correctly implemented, relative grading establishes a setting for cut-throat competition among students with no incentive for cooperation (which is
as much a part of life as is competition). The system also has absurd implications when the achievement of a group is for various reasons high or low, or when the group does not possess randomly distributed ability, motivation, and so on. Even proponents of normative-referenced evaluation implicitly accept the need for absolute grading, in that their reasons for the former approach are based on the idea that the stability of achievement within a relatively large group is supposedly a more accurate basis for grading than are an individual instructor's attempts to project absolute achievement.

If such non-relative judgments are possible, in that groups are sufficiently large and constant in terms of ability and other factors, so that such terms as sixty percentile do indicate a specified level of achievement on a consistent basis through time, then such normative-reference grading should be acceptable. But where such conditions do not exist, the criterion-reference system is preferable by far, and is more consistent with evaluation theory.

Obviously the necessary conditions for normative-referenced grading do not exist in relatively small groups, in classes where instruction is individualized in any way, or in those where students have variable time limits in meeting instructional objectives. Yet even where conditions favorable to normative evaluations do exist, so that normative- and criterion-reference grading may in practice be the same, it is probably more helpful to students to understand what level of knowledge and understanding of course material is required of them for a satisfactory grade, rather than to rely on a relatively content-less knowledge that seventy or some percentile will result in such a grade. This recommendation is most consistent with the high ratings of the feedback and motivation functions, and with the low ratings of selection functions obtained in the study, the former being more compatible with criterion reference, and the latter with norm reference.

Mastery-Based Instruction. Mastery-based instruction presumes that every or almost every student can master carefully defined course objectives and can also perform at least to some lesser minimally satisfactory level, providing that instruction is sufficiently individualized to the learning style of the individual and that sufficient time is
allowed to each student to achieve mastery. Such instruction also presumes that this possibility is more than theoretical, that at the present it can be implemented with reasonable resources. Since it is oriented to the success of the student, rather than to the relegation of almost every student to varying degrees of unsuccess, this approach is especially desirable. It coincides most closely, too, with the purpose of education, which is academic achievement unconditioned by other qualifications, such as achievement by this or that method within so many days and weeks. Understandably, the mastery approach has commanded increasing attention in recent years. Its influence on grading practices should be positive and it also should prove especially consistent with the strong concern of students for those grading objectives related to the teaching-learning process.

The Individual Institution. Each institution should formulate and enforce a policy on grading as well as an appropriate grading system. This would preclude the idea that academic freedom should be interpreted as the unlimited right of an individual instructor to teach the content he chooses, in the way he chooses, with his own particular method for evaluation and grading. Since in the case of evaluation and grading as well as in other areas, the actions of an individual instructor have institution-wide implications, divergent grading standards affect the overall academic program, enrollment patterns, and the quality of the graduate. The faculty of an institution, therefore, should establish a coherent grading program and be expected to conform to it. That policy and practice should be made known to students and also to external publics which use these grades.

Development of such a policy and implementing it in practice are difficult tasks. In order to accomplish them workshops, grading literature, and various learning experiences relating to evaluation and grading need to be made available to the faculty and administrators of the academic program. At the same time, specific procedures for monitoring the implementation of the policy should be developed, along with measures for enforcing its implementation.

Objectives of Grading Systems. There is at least one grading objective which ought to be abandoned, namely, selection
for employment opportunities. The prediction ability of grades in this regard is so low that it ought not to be considered ethical for colleges and universities to provide this information to prospective employers. While general intellectual ability presumably contributes to high grades as well as to job success, that ability should be directly measured, rather than estimated through grades. Grades are also affected by other factors, some of which, like conformity, for example, may be indicative of little success in a job. It follows, then, that grades and job success should not be considered related.

Selection for graduate and professional schools, however, is an appropriate objective. Assuming that they generally coincide with student achievement, grades may also be appropriately used for the granting of credit, honors, continuation, and other intra-institutional purposes. Grading should be defined broadly, to include the entire evaluation process rather than just the final bestowal of a letter or numerical symbol, so grades should be designed to provide as much information of a feedback nature as possible to faculty and student. That broad definition of grading is justified, since all evaluation would tend to be inadequate and haphazard unless formal product evaluation were provided.

This study indicates that students consider the objectives of feedback to instructor and student favorably, and that they at least accept the objectives of grades as recognition of achievement and selection tool for the school, but reject the use of grades for employment purposes or for the enforcement of non-academic achievement behavior standards, such as attendance and conduct. In this, student judgment is essentially sound, though the use of grades for selection by schools might deserve a more than acceptable rating.

Determinants of Grades. The only determinant of a grade should be student achievement of relevant educational objectives in a course. Although the literature of grading indicates that many other factors are intentionally or unintentionally used, there is no acceptable grading system which would allow a variety of miscellaneous factors to lead to the assignment of the same grade to students with varying levels of achievement. In some instances, of course, it may be pertinent to consider certain factors not normally
included under academic achievement, if they are demonstrably related to the educational objectives of a course. Creativity and imagination, for example, may be direct grading criteria in a fine arts course, but not in a medical technology course, and attendance might be considered in computing the clinical nursing grade, since professional responsibility is a specific goal for such courses, but it should not be considered in other courses except indirectly, in that non-attendance results in failure to achieve the direct course objectives. Neither should improvement during a course or any other such factor constitute criteria for assigning a grade. The one grading objective specifically requiring the use of such non-academic criteria was rated the lowest by students.

Characteristics or Criteria of a Satisfactory Grading System.
While many of the specific elements of a satisfactory grading system must wait for further research and study, some general considerations can be listed. The grading system must communicate accurately the nature of student achievement to all concerned parties, to the student, the instructor, and the users of grades. Grades assigned should reflect achievement of educational objectives only, not of unrelated factors. The grading system must, at minimum, not distort or hinder the teaching-learning function, and ideally should enhance it. In this regard the product evaluation required for grading should complement and reinforce the process evaluation which is essential for learning. To do this, the grade letters, symbols, or descriptions must be sufficiently simple to be administered, used, and understood and they must also be reliable and valid for their intended purposes. At least some form of criterion reference should therefore be incorporated into the grading system.

Specific Grading Systems. Although students, along with many others, react in stereotype fashion to a "yes or no" query about grades, this study, using an indirect approach, indicates that their response to several grading systems is distinctly positive. Grading systems which provide more precise and detailed information about academic achievement are preferred by students, and especially by more able students. On the basis of the reactions of the groups used in the study, there appears to be little need for individualized grading systems, though this may vary for more
specialized groups than were included. The study, like the literature, suggests that pass-fail grading need not be strongly considered, for contrary to most student opinion surveys, students in this case did not consider pass-fail grading especially effective in achieving grading objectives. Non-punitive grading which eliminates the D and F, a fairly recent innovation, though it seems to be consistent with educational objectives, likewise is not rated especially high by students. It is evident that a more descriptive and precise grading scheme than traditional grading is desirable.

Grading Dysfunctions. While no grading system can be developed simply for the sake of avoiding dysfunctions, these concerns recorded in the literature must certainly be granted consideration in any grading reform. The tendency to make grades rather than educational ends primary can occur only when the achievement of a particular grade does not necessarily coincide with the achievement of a particular level of an educational objective. Likewise, the frequently mentioned incentive for students to memorize trivia results from evaluation techniques which measure trivia rather than important educational objectives. True, the recognition of a variety of such grading dysfunctions cannot provide the rational for the ideal grading system, yet it can suggest what is to be avoided. As a general rule, no grading system will be satisfactory if it is not carefully implemented with sound evaluation techniques. The converse is also true; with careful evaluation and implementation, almost any grading system can be at least minimally satisfactory.

Prognosis. The prognosis for grading reform is both optimistic and pessimistic. On a large scale, grading has changed very little through the decades, the two most striking innovations having been rising grade distributions and the introduction of the limited pass-fail option. For reasons cited above, neither of these can be considered salutary. Furthermore, the quantity, though not the quality, of articles on grading has risen, but little substantive progress has resulted from such studies, though some developments portend hope. Grade inflation, for example, though not positive in itself, must eventually provoke a serious, thoughtful, and productive (?) reaction, since it is obvious that continued grade inflation would result in the abolition
of grading. On the brighter side, it must be noted that developments in systems and in evaluation theory are providing a better conceptual basis for considering grading. Besides this, since they do not fit well with traditional grading and grading theory, such innovations as credit by examination, credit for life experience, and mastery learning, are forcing a re-examination of grading. Hopefully, educational leaders will soon realize fully the tremendous implication of grades on learning and teaching and insist upon the comprehensive approach to grading reform which is required.

Grading systems, patterns, and attitudes are all deeply intertwined in the culture of individual institutions and higher education, but what has developed is in many ways irrational, not well understood, and dysfunctional although some important and valid purposes are served by grades.

A way of life, a culture, has developed around grading, a culture which means different things to different departments, instructors, and institutions, though working relationships at least have admittedly been forged regarding it. Any change in that system or in the status quo in the direction of grading reform will undoubtedly generate strong resistance, since such change also will undoubtedly effect basic attitudes and values for student, faculty, and administrator alike. Without strong and continuing incentive toward grading reform on a national basis, it is probable that significant change will occur only at a scattered and token pace at a small number of institutions--as has been the case to date.
LIST OF REFERENCES


Beyer, Harold N. "Effect of Students' Knowledge of Their Predicted Grade Point Average on Academic Achievement" *Journal of Counselling Psychology,* XVIII(Nov., 1971), pp. 603-605.


Blaylock, Mabry G. "Student Unrest From a Middle Ground." *Improving College and University Teaching,* XIX (Summer, 1971), pp. 211-213.


Chronister, Jay L. "Instructional Accountability in Higher Education." Educational Record, LII(Spring, 1971), pp. 171-175.


Claunch, Sidney J. "Effects of Pass-Fail Grading on Quality Point Averages." College and University, XLVII (Winter, 1972), pp. 93-105.


Committee on Educational Policy. "Report on Grading at the University of California, Santa Cruz." Santa Cruz, Calif.: University of California, 1970. (Mimeographed.)

Conway, J. A. "What Are We Rewarding?" Phi Delta Kappan, LI(Oct., 1969), pp. 87-89.


Ebel, Robert L. "Basic Considerations in Grading the Achievement of College Students." Lansing, Mich.: Michigan State University, Sept., 1966. (Mimeographed.)


Edison, Frank G. "Grade Distribution in Eighty Midwestern Liberal Arts Colleges." School and Society, LXXIII(May 19, 1951), p. 313.


Faust, Margaret A. In Opposition to Conventional Grades. Claremont, Calif.: Scripps College, Feb., 1971. ED 051 753.


Grade Distribution of College of Instruction. (Quarterly Report) Columbus, Ohio: The Ohio State University (Office of the Registrar), Autumn, 1956-1967 and Spring, 1968.


Greene, J. H., and Hicks, C. R. "Do College Class Grades Follow a Normal Distribution?" College and University, XXXVI(Spring, 1961), pp. 296-302.


Harris, J. W. "Baccalaureate Requirements: Attainments or Exposures?" Educational Record, LIII(1972), pp. 59-65.


Hauwiller, Robert. "Users Perceptions of Non-Traditional Grading Patterns: Undergraduate and Graduate Institutions and the Extensive Use of Highly Non-Traditional Grading Patterns--Pass/Fail or Credit/No Record." In a Report of the Subcommittee to survey the Acceptance of Non-Traditional Grading Patterns by Government, Industry and/or Graduate Institutions--1972. Edited by Robert L. Bailey. Park Forest South, Ill.: Office of Admissions and Records of Governors State University in Cooperation with AACRAO Committee on Institutional Studies and Operational Analysis, 1972.


Hayes, E. "Influence of Course Load on College GPA." College and University, XXXVII(Spring, 1962), pp. 251-253.


Hoyt, Donald P. "Rationality and the Grading Process." Educational Record, IL(Summer, 1970), pp. 305-309.


Kalsen, T. J. "Grades: Judgment or Lottery?" Improving College and University Teaching, XV(Summer, 1967), pp. 178-180.


"Law School Admission Test Council Statement on Pass-Fail Grading Systems as Endorsed by the Section of Legal Education and Admissions to the Bar of the American Bar Association." ED 045 043.


McHugh, L. "Repeat Course Policy." College and University, XLIV(Spring, 1969), pp. 279-282.


Marshall, Max S. "Triangular Grading." College and University, XLIII(1968), pp. 143-149.

Marshall, M. S. "Your Grades Are?" College and University, XLIV(Winter, 1969), pp. 182-188.


Miller, Richard H. "Students Show a Preparation Increase But No Increase in Grades Was Shown." *College and University*, XLV(Fall, 1969), pp. 28-30.


Moellenberg, Wayne. "To Grade or Not to Grade--Is That the Question?" *College and University*, IL(Fall, 1973), pp. 5-13.


Noble, Milton E. "Advantages and Disadvantages Associated With the 'No-Fail' Grading Systems." College and University, XLVI(Summer, 1971), pp. 717-726.


Page, Alex. "To Grade or Retrograde?" College English, XXI (Jan., 1960), pp. 213-216.


Pemberton, W. A. "The Grade Point Average: Snark or Boojum?" Newark, Delaware: Student Counseling Service, University of Delaware, Sept., 1970. ED 047 009.


Perry, Winona M. "Are Grades and Grading Systems Comparable From One Institution to Another?" Journal of American Association of College Registrars, XVIII(1943), pp. 159-165.


Quann, Charles J. "The Pass-Fail Option: Analysis of an Experiment in Grading." College and University, XLVI(Summer, 1971), pp. 542-549.


Reeves, Floyd W.; Peik, Wesley, E.; and Russell, John D. Instructional Problems in the University. Chicago: University of Chicago Press, 1933.


Scales, Eldridge E. "Variability of Grading Practices Among Instructors of a Multiple-Section English Course." College and University, XXXIII(Spring, 1958), pp. 334-336.
Scales, Eldridge E. "Effect of Instructor-Agreement on Evaluation Upon Assigned Grades in a Multiple Section Course." College and University, XXVI (Winter, 1961), pp. 201-204.


Shuman, R. B. "Does a C Mean Average?" Improving College and University Teaching, XIV(Spring, 1966), pp. 125-126.


Silvey, H. M. (Director) Study of the Grades Assigned at Iowa State Teachers College During the 1958-1959 Academic Year. Cedar Falls, Iowa: Iowa State Teachers College, Bureau of Research and Examination Services, 1959.


Statutes, By-Laws, Rules. Columbus, Ohio: The Ohio State University, 1968.


Swanson, Robert. Transferability of Graduate Work Graded According to Mastery: A Survey of Selected Graduate Faculty and Institutions. Menomie, Wis.: Stout State University, Graduate School, June, 1970. ED 049 712.


Taylor, Hugh. "Student Reaction to the Grade Contract."  


Thayer, Robert E. "Do Low Grades Cause College Students to Give Up?" Long Beach, Calif.: California State College, April, 1971. ED 054 725.


Thorndike, Robert L., ed. *Educational Measurement*  


Willingham, W. W. "Effect of Grading Variations Upon the Accuracy of Predicting Freshman Grades." College and University, XL(Winter, 1965), pp. 159-164.
Wilson, Kenneth M. "Increased Selectivity and Institutional Grading Standards." College and University, XLVI (Fall, 1970), pp. 46-53.


APPENDIX A

STUDENT GRADING QUESTIONNAIRE

COLLEGE OF EDUCATION
OHIO STATE UNIVERSITY

Freshman
Sophomore
Junior
Senior

Estimated Grade-Point Average: ________
Dear Student:

The value of grades in college, and in fact at all levels, has been increasingly questioned. Some persons, while granting that problems exist, maintain that grades are a necessary and integral part of higher education and do serve valid objectives. Others insist that grades are anti-educational and, whatever the purposes served, the results on balance are negative. Some new grading systems are being tried.

While many educators have expressed their opinions, student judgment has not yet been systematically studied. The attached questionnaire is designed to let you evaluate grades on the basis of your own experience. You are asked to judge the desirability of goals to be served by grading and how effectively you believe various grading systems might or do attain these goals.

Please answer each question carefully. Your evaluation is important.

Sincerely,

James F. Wasserman

Please indicate the course #________

Male______ Female______
Part 1

Rate the desirability—in your opinion—of the following goals to be attained through grading. Do not consider how well these goals are attained in practice. Circle the appropriate answer according to the scale:

0 - no desirability
2 - some desirability
4 - moderate desirability
6 - high desirability

A. Feedback, teaching tool for instructor
B. Feedback, learning tool for student
C. Motivation for student
D. Means of rewarding behavior desired by instructor (other than academic achievement—such as attendance, attitudes, etc.)
E. Mark of recognition for competencies or knowledge achieved
F. Selection/screening tool for college and graduate school. Basis for selection, awarding honors, retention in college, etc.
G. Selection tool for employer

Part 2

Rate the effectiveness—in your opinion—of each grading system described below according to its effectiveness in achieving the goals listed. Circle the appropriate answer
according to the scale:

<table>
<thead>
<tr>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

A. A, B, C, D, F with pass-fail option in one course per term, restricted to elective courses (4 point system)

1. Feedback, student                        0 1 2 3 4 5 6
2. Feedback, instructor                     0 1 2 3 4 5 6
3. Motivation, for student                  0 1 2 3 4 5 6
4. Selection, for college and graduate school 0 1 2 3 4 5 6
5. Selection, for employer                  0 1 2 3 4 5 6
6. Reward for desired behavior              0 1 2 3 4 5 6
7. Recognition, for achievement             0 1 2 3 4 5 6

B. A, B, C (grades of D and F are deleted from records and are not counted in grade average, no credit is earned)—grade-point-average is computed from A's, B's, and C's only, courses in which the student does not earn C or above can be repeated later for credit.

1. Feedback, student                        0 1 2 3 4 5 6
2. Feedback, instructor                     0 1 2 3 4 5 6
3. Motivation, for student                  0 1 2 3 4 5 6
4. Selection, for college and graduate school 0 1 2 3 4 5 6
5. Selection, for employer                  0 1 2 3 4 5 6
6. Reward for desired behavior              0 1 2 3 4 5 6
7. Recognition, for achievement             0 1 2 3 4 5 6
C. Descriptive grading. A brief, one paragraph evaluation by the instructor, such as: "Works hard, masters basic factual information. Shows some indications of creativity. Has difficulty in synthesizing and applying course material." The student's academic record is composed of the comments of instructors for all courses taken. No grade point average, etc. is computed; grades as such are not assigned.

1. Feedback, student
2. Feedback, instructor
3. Motivation, for student
4. Selection, for college and graduate school
5. Selection, for employer
6. Reward for desired behavior
7. Recognition, for achievement

D. A+, A, A-, B+, B, B-, C+, etc. (15 point system, +'s and -'s are counted in the grade-point-average)

1. Feedback, student
2. Feedback, instructor
3. Motivation, for student
4. Selection, for college and graduate school
5. Selection, for employer
6. Reward for desired behavior
7. Recognition for achievement
E. Pass-Fail system of grading for all courses (grade-point-average, etc. are not computed)

1. Feedback, student
2. Feedback, instructor
3. Motivation, for student
4. Selection, for college and graduate school
5. Selection, for employer
6. Reward for desired behavior
7. Recognition, for achievement

F. A, B, C, D, F (4 point system)

1. Feedback, student
2. Feedback, instructor
3. Motivation, for student
4. Selection, for college and graduate school
5. Selection, for employer
6. Reward for desired behavior
7. Recognition, for achievement

Part 3

Please record any comments you would like to make about grades or this questionnaire on the reverse side of this page. Thank you for your help.
<table>
<thead>
<tr>
<th>Grading Objective</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feedback for Instructor</td>
<td>4.00</td>
<td>1.68</td>
</tr>
<tr>
<td>Feedback for Student</td>
<td>3.75</td>
<td>1.49</td>
</tr>
<tr>
<td>Motivation for Student</td>
<td>3.82</td>
<td>1.66</td>
</tr>
<tr>
<td>Reward for Desired Behavior</td>
<td>3.79</td>
<td>1.48</td>
</tr>
<tr>
<td>Mark of Recognition</td>
<td>3.50</td>
<td>1.37</td>
</tr>
<tr>
<td>Selection for School</td>
<td>1.82</td>
<td>1.79</td>
</tr>
<tr>
<td>Selection for Employer</td>
<td>3.36</td>
<td>1.79</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grading System</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-F with P-F</td>
<td>99.68</td>
<td>48.90</td>
</tr>
<tr>
<td>A-C without D or F</td>
<td>76.61</td>
<td>42.77</td>
</tr>
<tr>
<td>Descriptive Grading</td>
<td>91.25</td>
<td>47.57</td>
</tr>
<tr>
<td>A-F with + and -</td>
<td>93.25</td>
<td>45.03</td>
</tr>
<tr>
<td>Total P-F</td>
<td>104.43</td>
<td>46.86</td>
</tr>
<tr>
<td>A-F (4 point)</td>
<td>96.25</td>
<td>48.47</td>
</tr>
</tbody>
</table>
APPENDIX B

TABLE 17

Design Layout for Student Rating Scores
for Importance of Grading Objectives

<table>
<thead>
<tr>
<th>S1</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
<th>B4</th>
<th>B5</th>
<th>B6</th>
<th>B7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>X1,1,1</td>
<td>X1,2,1</td>
<td>X1,3,1</td>
<td>X1,4,1</td>
<td>X1,5,1</td>
<td>X1,6,1</td>
</tr>
<tr>
<td>S100</td>
<td></td>
<td>X1,1,100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S101</td>
<td></td>
<td>X2,1,101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S200</td>
<td></td>
<td>X2,1,200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S201</td>
<td></td>
<td>X3,1,201</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S300</td>
<td></td>
<td>X3,1,300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A1 - Underclassmen
A2 - Upperclassmen - Arts and Sciences
A3 - Upperclassmen - Professional Areas
B1, ..., B7 - Objectives
S1, ..., S300 - Subjects
### Table 18

Design Layout for Student Rating Composite Summary
Scores for Effectiveness of Grading Systems

<table>
<thead>
<tr>
<th></th>
<th>C₁</th>
<th>C₂</th>
<th>C₃</th>
<th>C₄</th>
<th>C₅</th>
<th>C₆</th>
</tr>
</thead>
<tbody>
<tr>
<td>S₁</td>
<td>X₁,1,1</td>
<td>X₁,2,1</td>
<td>X₁,3,1</td>
<td>X₁,4,1</td>
<td>X₁,5,1</td>
<td>X₁,6,1</td>
</tr>
<tr>
<td>A₁</td>
<td>X₁,1,100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₁₀₀</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₁₀¹</td>
<td>X₂,1,101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₂₀₀</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₂₀¹</td>
<td>X₃,1,201</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S₃₀₀</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A₁ - Underclassmen  
A₂ - Upperclassmen - Arts and Sciences  
A₃ - Upperclassmen - Professional Areas  

C₁, ..., C₆ - Grading Systems  
S₁, ..., S₃₀₀ - Subjects
### TABLE 19

**Group Means for Rating of Objectives**

|       | B₁  | B₂  | B₃  | B₄  | B₅  | B₆  | B₇  | \(\bar{X}_A\) |
|-------|-----|-----|-----|-----|-----|-----|-----|              |
| A₁    | 3.70| 3.72| 3.92| 2.40| 3.40| 3.34| 3.00| 3.35         |
| A₂    | 3.65| 3.59| 3.37| 2.36| 2.93| 3.02| 2.51| 3.06         |
| A₃    | 3.79| 3.65| 3.40| 2.24| 2.86| 2.90| 2.33| 3.02         |
| \(\bar{X}_B\) | 3.71| 3.65| 3.56| 2.33| 3.06| 3.09| 2.61| \(\bar{X}_X = 3.14\) |

- **A₁** - Underclassmen
- **A₂** - Upperclassmen - Arts and Sciences
- **A₃** - Upperclassmen - Professional Areas
- **B₁** - Feedback for Instructor
- **B₂** - Feedback for Student
- **B₃** - Motivation for Student
- **B₄** - Reward for Desired Behavior
- **B₅** - Mark of Recognition
- **B₆** - Selection for School
- **B₇** - Selection for Employer
<table>
<thead>
<tr>
<th></th>
<th>C₁</th>
<th>C₂</th>
<th>C₃</th>
<th>C₄</th>
<th>C₅</th>
<th>C₆</th>
<th>Xₐ</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₁</td>
<td>85.01</td>
<td>81.81</td>
<td>92.70</td>
<td>90.30</td>
<td>47.87</td>
<td>85.25</td>
<td>80.49</td>
</tr>
<tr>
<td>A₂</td>
<td>65.77</td>
<td>73.71</td>
<td>90.09</td>
<td>79.72</td>
<td>46.84</td>
<td>68.34</td>
<td>70.75</td>
</tr>
<tr>
<td>A₃</td>
<td>71.14</td>
<td>66.21</td>
<td>82.22</td>
<td>75.08</td>
<td>49.78</td>
<td>70.65</td>
<td>69.18</td>
</tr>
<tr>
<td>Xₐ</td>
<td>73.98</td>
<td>73.91</td>
<td>88.34</td>
<td>81.70</td>
<td>48.76</td>
<td>74.75</td>
<td>X = 73.47</td>
</tr>
</tbody>
</table>

A₁ - Underclassmen
A₂ - Upperclassmen - Arts and Sciences
A₃ - Upperclassmen - Professional Areas

C₁ - A-F with P-F
C₂ - A-C without D or F
C₃ - Descriptive Grading
C₄ - A-F with + and -
C₅ - Total P-F
C₆ - A-F (4 point)
TABLE 21
Mean Ratings of Grading Objectives by Students
According to Grade Point Average Intervals

<table>
<thead>
<tr>
<th>GPA Interval</th>
<th>B₁</th>
<th>B₂</th>
<th>B₃</th>
<th>B₄</th>
<th>B₅</th>
<th>B₆</th>
<th>B₇</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00-1.49</td>
<td>5.00</td>
<td>6.00</td>
<td>4.00</td>
<td>6.00</td>
<td>2.00</td>
<td>2.00</td>
<td>6.00</td>
<td>1</td>
</tr>
<tr>
<td>1.50-1.99</td>
<td>0.00</td>
<td>0.00</td>
<td>2.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.00</td>
<td>1</td>
</tr>
<tr>
<td>2.00-2.49</td>
<td>3.68</td>
<td>3.44</td>
<td>3.23</td>
<td>2.39</td>
<td>2.61</td>
<td>2.47</td>
<td>2.23</td>
<td>57</td>
</tr>
<tr>
<td>2.50-2.99</td>
<td>3.61</td>
<td>3.64</td>
<td>3.75</td>
<td>2.21</td>
<td>3.14</td>
<td>2.98</td>
<td>2.37</td>
<td>89</td>
</tr>
<tr>
<td>3.00-3.49</td>
<td>3.94</td>
<td>3.68</td>
<td>3.61</td>
<td>2.52</td>
<td>3.19</td>
<td>3.44</td>
<td>2.88</td>
<td>69</td>
</tr>
<tr>
<td>3.50-4.00</td>
<td>4.32</td>
<td>4.12</td>
<td>3.44</td>
<td>1.94</td>
<td>3.12</td>
<td>3.29</td>
<td>2.88</td>
<td>34</td>
</tr>
<tr>
<td>Total</td>
<td>3.81</td>
<td>3.67</td>
<td>3.55</td>
<td>2.31</td>
<td>3.01</td>
<td>3.02</td>
<td>2.56</td>
<td>251</td>
</tr>
</tbody>
</table>

B₁ - Feedback for Instructor
B₂ - Feedback for Student
B₃ - Motivation for Student
B₄ - Reward for Desired Behavior
B₅ - Mark of Recognition
B₆ - Selection for School
B₇ - Selection for Employer
### TABLE 22

Mean Ratings of Grading Systems by Students According to Grade Point Average Intervals

<table>
<thead>
<tr>
<th>GPA Interval</th>
<th>C₁</th>
<th>C₂</th>
<th>C₃</th>
<th>C₄</th>
<th>C₅</th>
<th>C₆</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00-1.49</td>
<td>128.00</td>
<td>121.00</td>
<td>130.00</td>
<td>106.00</td>
<td>112.00</td>
<td>118.00</td>
<td>1</td>
</tr>
<tr>
<td>1.50-1.99</td>
<td>0.00</td>
<td>8.00</td>
<td>0.00</td>
<td>8.00</td>
<td>4.00</td>
<td>8.00</td>
<td>1</td>
</tr>
<tr>
<td>2.00-2.49</td>
<td>61.26</td>
<td>69.93</td>
<td>79.63</td>
<td>72.75</td>
<td>45.68</td>
<td>67.46</td>
<td>57</td>
</tr>
<tr>
<td>2.50-2.99</td>
<td>73.89</td>
<td>75.52</td>
<td>87.02</td>
<td>77.89</td>
<td>51.96</td>
<td>69.67</td>
<td>89</td>
</tr>
<tr>
<td>3.00-3.49</td>
<td>73.91</td>
<td>71.93</td>
<td>94.59</td>
<td>83.93</td>
<td>45.15</td>
<td>78.17</td>
<td>69</td>
</tr>
<tr>
<td>3.50-4.00</td>
<td>81.32</td>
<td>70.94</td>
<td>95.53</td>
<td>97.62</td>
<td>46.06</td>
<td>84.50</td>
<td>34</td>
</tr>
<tr>
<td>Total</td>
<td>71.96</td>
<td>88.40</td>
<td>88.40</td>
<td>80.89</td>
<td>47.91</td>
<td>73.46</td>
<td>251</td>
</tr>
</tbody>
</table>

- C₁ - A-F with P-F
- C₂ - A-C without D or F
- C₃ - Descriptive Grading
- C₄ - A-F with + and -
- C₅ - Total P-F
- C₆ - A-F (4 point)