The traditional A-B-C-D-F grading system utilized by most colleges and universities to indicate student achievement has been a major focus of student and faculty dissatisfaction for decades. Further, when connected (consciously or unconsciously) to a norm-referenced approach to testing and a curve-based marking practice, the system can and does work serious injustices, requiring that some students "fail" and others be scored "average" in spite of their remarkable achievements and the fact that they have learned the material. These problems have not gone unnoticed. Beginning in the 1940's, increasing numbers of institutions at all levels of higher education began to experiment with innovative grading systems. This bibliography is not the first survey of the field of literature relating to innovative grading practices. Rather, the purpose of the present bibliography has been to update foregoing efforts and tap research sources that appear to have been neglected. It focuses on grading innovations at all levels of undergraduate higher education: universities, colleges, and community/junior colleges. Some valuable insights from secondary school studies have occasionally been included. A total of fifty-one reports, documents, and dissertations are included in this annotated bibliography. (Author/JDS)
INNOVATIVE GRADING PRACTICES
AN ANNOTATED BIBLIOGRAPHY

Terry H. Smith Wallace
Associate Professor
Harrisburg Area Community College
Harrisburg, PA 17011
Introduction

The traditional A-B-C-D-F grading system utilized by most colleges and universities to indicate student achievement has been a major focus of student and faculty dissatisfaction literally for decades. Further, when connected (consciously or unconsciously) to a norm-referenced approach to testing and a curve-based marking practice, the system can and does work serious injustices, requiring that some students "fail" and others be scored "average" in spite of their remarkable achievements and the fact that they have learned the material. The irony in the utilization of the 'bell-shaped' curve -- that statistical reflection of random distribution due to random activity -- is harsh in an era when higher education is under attack to prove its productivity, for curve-based grading practices at worst suggest education to be random activity reflected in a random distribution of achievement scores, or at best a rather cruel, archaic, and ineffective people sorting system (Hoyt, 1966) -- one which appears to be extraordinarily expensive in terms of both wasted money and wasted human potential.

Clearly, these problems have not gone unnoticed or untouched. Starring in the meritocratically oriented 1940s and 1950s, and gaining momentum under the egalitarian outlook and campus unrest of the 1960s, increasing numbers of institutions at all levels of higher education began to experiment (some timidly, others radically) with innovative grading systems. This trend shows little sign of abating in the late 1970s, receiving new emphasis by the demand on higher education to substantiate its claims, to be accountable for student learning, and to move in the direction of competency-based performance oriented education.

With such a significant growth in experimentation with grading systems, it is not surprising that the literature describing such efforts has expanded with extraordinary rapidity. Thus, the present bibliography is not the first survey of the field. Caldwell (1973) has done an annotated review of research on grading systems of possible significance to junior colleges, while Quann (1970) and Collins and Nickel (1974) have reported major surveys of grading practices and trends at institutions of higher education. Moreover, specific marking systems which appear to be particularly popular, like pass-fail grading and computer grading, have received major bibliographic attention (Davidovicz, 1972; Pedrini and Pedrini, 1972; Levy and Fritz, 1972).

The purpose of the present bibliography has been to update foregoing efforts and tap research sources they appear to have neglected. It focused on grading innovations at all levels of undergraduate higher education: universities, colleges, and community-junior colleges. However, unusually valuable insights from secondary school studies have occasionally been included, especially when they have ventured into new areas as yet untried by those in higher education, or when such work was deemed
corroborative of similar experiments in colleges and universities. Further, a definite attempt was made to avoid duplicating the efforts of those bibliographies mentioned above. The aim here was to update their efforts and supplement them, the latter by tapping research document sources which the foregoing compilers have seriously neglected, sources of valuable but low visibility studies. Thus, dissertations and ERIC documents have of necessity received a disproportionate amount of attention in the present work.

Writing in Dateline '79: Heretical Concepts for the Community College, Arthur M. Cohen (1969, pp. 129-130) predicted that in the community college of 1979 there would be no grading-marking and no norm-referenced tests. When a student demonstrated his achievement of course objectives, he would be certified as competent, as able to perform the tasks for which the curriculum trained him. Obviously, competency-based performance-oriented grading systems have not swept through the community college movement like wild-fire (let alone the rest of higher education) and will not by 1979. Such fundamental changes in higher education rarely come so quickly. However, if a new trend is discernible in the extensive literature now available on grading innovations, it is that over the last five years institutions have begun in increasing numbers to move away from rather superficial attempts to solve the grading dilemma (e.g., pass-fail, credit-no credit) and to move towards in depth experiments. Those in depth experiments are ones of vital significance for they attack the core of the problem, the questions with which generations of students and faculty have wrestled: What are the purposes of grades? Do they support or hinder an educational institution's attempt to achieve its goals? What do grades mean in terms of learning, of student performance and competency? How well do they measure that learning? With the growth of significant efforts to answer these questions, the trend will most likely shift increasingly toward criterion-referenced competency-based performance-oriented marking systems. While Cohen's vision of grading in 1979 may be slower in coming than he surmised, its outlines are beginning to take shape in reality.

Terry H. Smith Wallace
10 August 1976
ERIC DOCUMENT (ED) numbers and University Microfilm order numbers are presented, whenever possible, to expedite the retrieval of information, especially of items not easily obtainable from other sources.

The AACRAO Survey of Grading Policies in Member Institutions.  

Proposed to delineate the nature and extent of changes from the traditional grading system, practices in accepting transfer credits from colleges with innovative grading systems, the rate and recency of change and the anticipated nature of systems in the near future. Many institutions indicated they are mixing innovative with traditional grading systems. However, responses to the survey suggested major changes in systems were accelerating.

Barthlow, R. L. Grade Point Average and Academic Probation in the Two-Year College.  
(ED 095 967)

Suggests a change from the fixed GPA standard used at many community colleges when determining academic standing to a developmental GPA standard for determining probation and dismissal. Claims such a system would allow the non-traditional student time to build himself up academically and that the technique would be in line with the open-door policy of community colleges. Gives no data to support system.

Bils, J. A. "The Relationship of Particular Secondary School Grading Systems to the Nature of the Organizational Systems and the presence of Individualization of Instruction."  
Ann Arbor, Mich.: University Microfilms (Order No. 73-30,538)

Investigated 2 questions: Are certain grading systems associated with more "open" or "closed" organizational systems? And, do certain grading systems facilitate or retard individualized instruction? Grading systems with high differentiation and little personalized information were considered school-centered, while systems with the opposite characteristics were classified as student-centered. Discovered that student-centered grading systems are associated with energetic, open organizational systems which, in turn, correlated to greater individualization of instruction.

Discusses the necessity (because instructors are expected to assign grades) of differentiating between the A, B, C, D and F levels of achievement when writing instructional objectives. Gives examples of multiple forms of the same objective written at different levels of achievement.


A first section presents the annotated bibliography; a second describes grading policies being utilized by 9 different community colleges.


Provides information on the current grading, recording, and averaging practices of 544 institutions of higher education. Findings suggest: (1) institutions are experimenting with a wide variety of practices, (2) the grading systems studied were very diverse in nature, (3) most colleges use one or two types of non-traditional grades, (4) the trend is toward less punitive grading in higher education, and (5) as competency-based education grows, the traditional transcript GPA will lose its historical meaning.


Attempts a comprehensive compilation of factors that could cause a student to fail in a given endeavor. Reviews factors in the areas of basic academic skills, work habits, career decisions, special skills, school, home, community, and peer group. Does so in hopes that educational program planners will consider these factors and attempt to avoid program elements that may lead to, or enhance them.


Reviews a 1968-1971 literature sample and finds that (1) student performance declines both in pass-fail courses and traditionally
evaluated courses once the student has taken the former, (2) students take pass-fail courses in order to ease their course work, (3) first-year students suffer most from pass-fail grading, (4) the elimination of failing grades from transcripts is increasing, (5) most institutions avoid problems with pass-fail by limiting the option, (6) students graded on pass-fail are less likely to get into the graduate school of their choice.


Investigated the effects of a nonpunitive A-B-C grading system and a traditional A-B-C-D-F system on student anxiety, student grading preference, and academic preference. Found (1) neither system had a relationship to student anxiety or vice versa; (2) when given a choice, students were inclined to prefer the grading system they were most familiar with— they did not necessarily prefer a non-punitive over a traditional one; and (3) the non-punitive system did not increase academic performance.

Ebersole, J. F. A Study of the "In-Progress" Grade. Practicum, Nova University, March 1975. (ED 104 452)

Examines the use of the "In-Progress" (Y) Grade as added to the generally traditional grading policy of Harrisburg Area Community College (PA) in November 1971. The "Y" grade gives the student the option of repeating a course, not successfully completed the first time around, to obtain a passing grade. The study reviews the use of the "Y" grade over 3 1/2 years and whether students have benefited from its use. Reports the device as moderately successful.


Reviews the historical backgrounds of traditional and non-punitive grading, the national trends toward non-punitive grading innovations, the legal background of the grading issue (including local California regulations), and an analysis of the grading standards and practices of the 5 Peralta colleges.

Discusses the variety of grading approaches, norm referenced and criterion referenced systems, the not always clear relationship between grades and learning, the problems involved in determining grades and the advantages and disadvantages of pass-fail and non-punitive systems. Suggests the grading dilemma will not be quickly solved.


Findings suggest "that the system of having Freshman students.....grade/edit each other's themes in English Composition.....can cause significantly greater gains in their grammar-usage ability as well as their theme composition ability than students whose themes were graded/edited by the course instructor."


Attempted to measure the use of an encounter tape experience (from the theoretical base of Rogers' [1969] support of the encounter experience) with college instructors with regard to grading practices. Discovered all members of the experimental group showed an increase in the grades they assigned. Results were considered beneficial and the program easily implemented at orientation sessions at the beginning of the year.

Concluded that pass/fail grading does not increase student achievement, enhance attitude, or reduce anxiety.


Describes an experiment (at SUNY, Cortland) to determine whether or not academic achievement is enhanced under pass-fail grading. Found the achievement of first-year students who took all of their courses on a pass-fail basis significantly lower than for controls who wanted but were denied pass-fail grading. Upper classmen allowed to take one course so graded were also significantly lower in achievement than controls who were denied the option. Further, "there was no compensatory improvement in the grades received in non-pass-fail courses."


Reports on an experiment with low achieving eleventh and twelfth graders in which students completing all assignments were insured against failing grades. Discovered that students in the Treatment group finished more homework assignments and had less than half the failures of the Control group. Found students who were offered no-fail insurance achieved at a higher level than the control group on three post-tests, homework, classwork, notebook, extra credit work, attendance, and motivation.


Discusses the spiraling GPA, traditional grading and alternatives, an honors/credit/no record option, and competency-based evaluation.


Examined student self-evaluation by collecting, compiling and examining the philosophies motivating, and the results of, the various programs utilizing it. Found (1) evaluation of high achievers more closely matched teacher judgments than those of lower achievers; (2) evaluation was ineffective when
competition was emphasized, but more effective when cooperation was stressed; (3) a general trend for students to show higher achievement under self-evaluation than under teacher evaluation; (4) the acceptance of student evaluation by those who 'consume' school evaluations appeared to be a major obstacle to its acceptance as an alternative to the traditional system.


Compared "an individualized instruction-contract grading program with a lecture-discussion section in a College Algebra and Trigonometry class with respect to achievement and attitude." Found no significant difference between the two groups in either achievement or attitude. The work also analyzes the advantages and disadvantages of the individualized instruction - contract grading program.


Proposed "to examine the relationship of students' and instructors' evaluations, to discover which elements of evaluation seem to best predict the grade awarded by each group, and to compare the dimensions of evaluative judgments for students and instructors." Found both constituencies appear to come to very similar conclusions on speeches, but for different reasons.


Sought "to determine if a qualitative and/or quantitative difference in cognitive performance proficiencies results from evaluating and rewarding college students as autonomous learners as compared to evaluating and rewarding them as intradependent team members." Found that (1) evaluation of students as intradependent team members does not significantly affect their achievement at various cognitive levels; (2) students prefer stated objectives to direct their study; (3) students will spend more time, working independently within a traditional evaluation system, than within the same system as intradependent team members; (4) most students consider grading of students as members of an intradependent team to be unfair.

Suggests "that college grades have no more than a very modest correlation with adult success, no matter how defined" and discusses 3 major implications of this finding. First, it recommends that the true meaning of grades be clearly established—that researchers delineate what they do measure. Second, it suggests major alterations in the evaluation system to encourage instructors to grade on a variety of considerations, not only knowledge, and to replace the traditional grade transcript with a more comprehensive profile of student development. Finally, it notes that the primary use of grades as a criteria to admission for professional training is indefensible, and that better indicators of professional potential should be sought.


Summarizes policies that normally govern pass-fail grading at most colleges and universities, and notes some tentative conclusions on student use of, and student achievement under the option. Suggests pass-fail be utilized chiefly for grading general education courses.


Found low utilization of the Satisfactory-Unsatisfactory (S-U) grading option. Discovered students usually used S-U system to avoid worry about a grade or to protect their GPA. Students reported more positive than negative aspects of the S-U option.


Suggests traditional forms of grading are incompatible with individualized instruction and that, in place of frequent feedback mechanisms and traditional grading, a testing service be established "that would occasionally examine, rank, and report on students' academic ability." Feels concept is particularly well suited to individualized instruction.

Reviews literature on computer utilization to grade English compositions. Discusses attempts to grade content as well as style.


Reports on elimination of plus and minus grades, the abolition of the "F" grade, and the liberalization of course withdrawal. Using "before" and "after" reform samples, the study found the new system had no significant effect in changing GPAs or the proportion of penalty and withdrawal grades to non-penalty grades. A significant difference was found in the persistence in college rate: the non-punitive grading system adversely affecting student registration for the second term (however, author reports threats to internal validity of study which call into question this last conclusion).


Found no significant difference between the mean GPA of pass-fail enrollees and a representative sample of regular enrollees. The pass/fail system seemed to be motivating students to take courses they would not otherwise have taken, but only a minority suggested it encouraged them to work harder.


Found that students view their pass/fail courses less seriously than their other courses, that their achievement levels tend to decline in pass/fail courses. Student reasons for selecting the pass/fail option were those commonly noted as its purposes and objectives.


Respondents noted relative acceptance of various systems. Also describes kinds of supplementary information desired by some, and how academic records are treated. Returns were seriously deficient
from graduate departments and employees surveyed, thus damaging validity of some of the conclusions.

McDonald, W. V. What Do We Do When We Do What We Do - And Should We? Paper presented at the Annual Meeting of the Conference on College Composition and Communication, April 1974. (ED 090 565)

Grading alone in composition can be misleading. Suggests that a written statement about the student's writing ability -- both strengths and weaknesses -- should be included in assigning a grade.


Investigated the effect of contract grading on student performance (final course grades, attitudes toward course, amount of time devoted to course, readings completed, and student perceptions of how much course fulfilled their needs). Discovered no significant differences on any of the foregoing measures of performance (except final course grades, and those only in a second semester replication).


Alberta's experience and pass/fail literature indicate the system does not appear to motivate learning, does not seem to motivate exploration outside of student's major, does seem to encourage laxness in performance.


Summarizes some of the issues raised by supporters and by critics of traditional grading, pass-fail grading, and a mastery learning model of grading. Suggests principles to which any system should adhere.


Available data references are reported concerning pass-fail techniques.

Summarizes a survey of grading options at selected 4-year colleges and universities, considering pass/fail, pass/no pass, satisfactory/unsatisfactory, credit/no credit, honors/pass/fail, pass/D or F, and 'other.' Pass/fail appeared the more prominent trend in new grading approaches, but the credit/no credit option—eliminating the idea of failure—seemed to be an emerging pattern, one to be watched.


The investigation involved 85 students divided into two comparable college courses (Educational Psychology) taught by the same instructor. Found no significant differences in achievement, attitudes, and values when a pass-fail system was substituted for a traditional ABCDF grading system.

Scully, M. G. "No Grades, No Credits, But 40 'Competency Units.'" The Chronicle of Higher Education, 3 February 1975, p. 5.

Describes the competency-based Liberal Arts program Alverno College (Wisconsin) developed in 1973. Liberal Arts competencies are defined in terms of processes the student can master and the college can measure. Students amass competency-level units indicating how well they can perform the tasks the college believes every graduate must possess.

Smith, L. G. Non-Punitive Grading in California Junior Colleges. 1969. (ED 035 409)

Briefly reports on a survey of Presidents and/or Deans of 94 California community colleges on what percentage of their faculty utilized a non-punitive A, B, C, W (withdrawal) grading system. Discovered growing interest in the system and increasing use. Discusses pros and cons of non-punitive grading.


Describes the A, B, C, I, WP, WI performance based grading system adopted by College of the Mainland. Gives the history of the system and its philosophical rationale. Explains how course planning and instructional objectives are integrated with the system and its implementation.

Describes some of the effects of the elimination of D and F grades at Flathead Valley Community College (Montana), University of Oregon and Grinnell College (Iowa). Discovered that (1) academic standards were neither weakened or strengthened by the change; (2) students preferred the non-punitive system, while faculty ranked the ABCDF system over it; (3) most respondents attributed no appreciable academic changes to non-punitive grading; (4) the contention that it was easier to earn C's with D's and F's eliminated was not supported by grade distribution; (5) though graduate school admissions chances were not negatively affected by the change to a non-punitive system, students were sometimes required to submit additional records at about 25% of the reporting institutions.


Describes an effort at Washington State University in 1968 "to assess student opinion in a senior educational measurement and evaluation course." Students viewed the contract system as a fairer one for assigning grades than conventional techniques. However, no significant differences were discovered between the treatment and control groups on the validity of the course examination, knowledge of the course objectives, or the quality of course standards.


Suggests few students use the option in their major study area, but many use it to satisfy 'outside college' requirements. Mean GPAs among credit-no credit users were consistently higher than the means of non-users in each class (except for freshmen). Suggests the option should be limited to upper-classmen and that efforts should be made to make it policy-wise a more attractive grading option.


Experiment with two treatment groups: one receiving letter grades for its composition and one receiving none. Both received identical treatment in other ways, including the use of positive comments on all papers. Found that in the presence of positive comments, the presence or absence of letter grades does not significantly affect change in attitude or writing performance as measured by study.

A major summary of literature up to 1971. Attempts a comprehensive review of the literature examining grading practices, uses of grades and their influence on the students, faculty and administration. Suggests that, in 1971, the liveliest issues in the debate were not the fundamental ones of what do grades represent? What are the purposes of grades? Are they meeting those purposes? Are there better ways to meet them? The concerns in 1971 were more superficial, oriented toward the number of grade categories, grade prediction, etc.


Analyzes the conditions of use of the credit/no-credit grading option. Found that student reasons for selecting the option were not significantly different from the basic objectives officially established for its use. However, mean grades achieved by students in credit/no-credit courses were significantly below the mean grades of those being traditionally graded in the same courses. Further, the option did not appear to be motivating students to take heavier course loads.


Studied a method for reducing the problem of scoring essay exams for subject matter by providing a computer with a scoring key developed from expert information. Attempted to ascertain which method would enable computer to examine content as competently as a human judge. Found that "intercorrelations of human judges with the computer were lower than those among human judges" -- in effect the human grader was better.


Study done at the Madison Consolidated Junior High School in 1972 found neither grading system fostered significantly greater learning, more positive learning, nor greater retention of knowledge learned. Concluded that the grade contract method is as effective as the traditional.

Questions use of grading as an evaluation of student performance. Suggests an alternative approach (Complete-Incomplete) to meet what authors believe are the proper criteria for evaluation:
(1) facilitation of self-evaluation, learning, and teaching,
(2) utilization of instructional objectives,
(3) provision for feedback on problems of curriculum development and educational policy,
(4) production of appropriate records.