2.5
2.8
2.2
3.2
2.0
3.6
4.0
100 mm
1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890
1.0 mm
1.5 mm
2.0 mm
A5

DOCUMENT RESUME

ED 279 713 TM 870 158

AUTHOR Stiggins, Richard J.; And Others
TITLE Inside High School Grading Practices. The Northwest Regional Educational Laboratory Program Report.
INSTITUTION Northwest Regional Educational Lab., Portland, Oreg.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 30 Nov 86
CONTRACT 400-86-0006
NOTE 37p.; Appended "Reasons for Discrepancies" pages contain small print.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Classroom Techniques; *Evaluation Methods; Grades (Scholastic); *Grading; *High Schools; Research Methodology; *Secondary School Teachers; Standards; *Student Evaluation; Teacher Attitudes; Teacher Improvement

ABSTRACT

This investigation was undertaken to provide insights as to how to improve the quality and relevance of teacher training in grading practices. The study was designed, by looking at grading practices of 15 high school teachers via intensive case study methodology, to explore: (1) the nature and technical quality of assessment and grading practices; and (2) why professional training has had so little impact. The researchers prepared a comprehensive framework of 34 grading issues to serve as a basis for observing teachers' grading practices. Information was gathered from each of the teachers in relation to the questions about grading practices. Seven questions focused on basic assumptions or antecedents that feed actual grading practices; twenty dealt with grading practices themselves; and the remainder addressed issues of the effects of grading. Discrepancies between best practices and actual practices were noted in 26 of the issues. An analysis of possible causes for the discrepancies revealed that 21 of the 26 practices probably have multiple causes. Recommended practice may need to be reevaluated in light of the realities of the classroom, and training in sound grading practices for teachers and principals is needed. (Results are presented for each issue, and reasons for discrepancies between recommended and actual practice are listed in chart form.) (LMO)

THE Northwest Regional Educational Laboratory

PROGRAM REPORT

INSIDE HIGH SCHOOL

GRADING PRACTICES

Richard J. Stiggins
Northwest Regional Educational Laboratory

Philip Griswold
Parkrose (OR) School District

and

David Frisbie
University of Iowa

November 30, 1986

Northwest Regional Educational Laboratory
300 S.W. Sixth Avenue
Portland, Oregon 97204

Sponsored by OERI Office of Educational Research and Improvement U.S. Department of Education

This publication is based on work sponsored wholly, or in part, by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number 400-86-0006. The content of this publication does not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. Government.

Abstract

The grading practices of 15 high school teachers were studied to determine the assumptions feeding into those practices, grading methods used and the effects of those methods. Thirty-four grading issues were addressed, comparing recommended and actual practices. Discrepancies were uncovered in 26 of the issues. Several possible reasons for these differences are explored, including unrealistic expectations in terms of recommended practice and a lack of teacher training in assessment and grading.

Introduction

Much of the professional literature related to grading practices in schools can be classified as (a) a statement of values regarding grades and grading, or (b) a description of available grading procedures. For example, Bail (1983) represents a value statement in which the author contends that grading systems should be evaluated in terms of the type of long-term/short-term learning that they reward. Current grading systems are said to reward short-term retention. Sadler (1983) contends that grading systems should encourage improvement by making performance expectations very clear to students. Ebel (1980) argues for testing and grading as a way to identify and reward excellence and to help students set realistic expectations. And Simon (1972) examines some of the key underlying assumptions of current grading practices (e.g., they prepare students for life in a competitive society) and calls these assumptions into question.

Some analyses of grading procedures reflect in a scholarly manner on the various grading functions and options. For instance, the National Education Association (1974) presents a thorough review of the historical evolution of grading practices. Simon and Bellanca (1976) review research on grading practices as that research relates to four key functions of grades: administrative, informational, motivational and guidance. Alternative grading methods are discussed. The measurement textbook by Hills (1981) presents a very thorough analysis of grading options, potential ingredients in determining a grade and technical problems to be addressed in grading.

Other analyses report the results of surveys of grading practices. Examples include a study of alternative grade reporting forms by Educational Research Service (1977) and a large-scale study of teachers' assessment and grading practices in Great Britain by Clough, Davis and Sumner (1984).

However, nowhere among these studies is there reported an indepth examination of the day-to-day practices teachers use to grade student performance. Nowhere is there an analysis of the underlying assumptions and philosophies teachers bring to the grading processes. Nowhere is there a summary of the actual practices teachers use to generate grades: the student characteristics they measure, the measurement procedures they use, their rules of evidence, the standards they apply. Rarely is there an inquiry as to the effects of grades on students and on their teachers (Cullen, 1975 does present one such study). The study reported here was intended to begin to fill these gaps in available knowledge about grading practices.

The investigation was undertaken to provide insights as to how to improve the quality and relevance of teacher training in grading practices. As the results reported below will indicate, teachers seldom cite preservice or inservice teacher training as a factor in determining the assessment or grading practices they use. The primary source of the knowledge upon which those practices are based is the teacher's experience as a teacher and before that as a student.

As a result, one must wonder about (a) the nature and technical quality of assessment and grading practices, and (b) why professional training has had so little impact. This study was designed to explore both of these issues by looking closely at the grading practices of fifteen high school teachers via intensive case study methodology. The researchers prepared a comprehensive

framework of 34 grading issues to serve as the basis for observing the teachers' grading practices. Each issue (e.g., What student characteristics do you consider in assigning grades?) was selected for inclusion in the study in part because the educational measurement community can specify recommended answers or best practices with regard to that issue. The objective of the study was to specify recommended practices, document actual practices and contrast the two. We sought to identify commonalities and discrepancies and to explain why the discrepancies might exist. Three possible explanations were considered:

1. Best practices may simply be a matter of opinion rather than technical standards. Discrepancies may reflect differences of opinion.

2. The measurement community may not understand the practical constraints or realities of the classroom that make recommended practice inappropriate.

3. Teachers may lack sufficient knowledge and skills to meet the recommended standard.

Such an analysis of classroom grading practices in relation to recommended practices and an examination of the reasons for discrepancies promised to reveal how teacher training in this important arena might be adjusted and made more relevant to and helpful for teachers.

Research Methodology

Information on grading practices was gathered via indepth case studies of a limited number of teachers. This method was selected for two reasons. First, this was an exploratory study designed in part to identify--not answer--the questions we must ask if we are to gain a clearer understanding of grading practices. Thus generalizability of results to a large population of

teachers was less important than gaining indepth information. In such a research context, case study methodology is advisable. Second, the objective of the study was to probe deeply into the underlying assumptions and methods of grading. This requires investment of resources in a high resolution research methodology capable of yielding the depth of understanding only achieved through case studies.

Teachers Studied. A total of 15 teachers were studied. All were high school teachers who taught a range of content in their subject matter area. Of this total, 4 taught math, 4 taught language arts, 5 taught science and 2 were social studies teachers. All were veteran teachers with from 5 to 35 years experience. Seven of the 15 case studies focused on grading practices only, while the remainder were part of a large study of classroom assessment and grading practices. All of the teachers volunteered to participate.

Issues Investigated. Information was gathered from each teacher in relation to 34 questions about their grading practices. Seven questions focused on basic assumptions or antecedents that feed into actual grading practices. Twenty dealt with grading practices themselves. And the remainder addressed issues of the effects of grading. The specific questions are listed in the results section which follows.

These issues were identified by the researchers based on (1) prior studies of classroom assessment (e.g, Stiggins and Bridgeford, 1985); (2) the researchers' technical expertise and experience in the field of educational measurement; (3) analyses of recommended grading practices presented in measurement textbooks such as those by Mehrens and Lehmann (1984); Ebel & Frisbie (1986), Hills (1981) and Gronlund (1981).

Data Collection. All participating teachers were observed in the classroom context. The frequency and length of observations varied from

teacher to teacher, depending on the researchers' level of understanding of the teacher's practices. Each teacher was interviewed, sometimes on several occasions, and most provided records and documentation of grading practices for review by the researchers.

Results

The results are presented below for each issue. Recommended practice is presented along with a synthesis of actual classroom practices across all cases. Discrepancies are identified and discussed briefly.

ANTECEDENTS

1. What is the purpose for grades?

Recommended Practice: Grades serve to communicate to parents, students and school staff the extent to which students have mastered the material taught. They also serve as aids to decision making about new material to be covered, promotion, placement and graduation. They serve to motivate student learning.

Case Results: Grades are intended to provide information to students and parents about student achievement at a particular point in time. While they rarely aid in decisions about new material to cover, grades do bear on placement, promotions and graduation decisions. They give students an incentive to do the required work.

Discrepancy: None.

2. What underlying principles guide grading practices?

Recommended Practice:

1. Grading practices must be clearly stated and public information.

2. Underlying measures must be valid and reliable.

3. Grades should reflect only the amount (or percent) of required content and skills the student has mastered (i.e., achievement).

4. Factors such as effort, attitude, and attendance should not be measured separately and factored into grades, as they are already reflected in the amount learned by student (achievement).

5. All students should have an equal opportunity to succeed and attain a high grade; a prior distribution of grades (i.e., grading on a curve) is not generally acceptable.

6. Grades should be objective, i.e., reproduceable by others using existing records.

Case Results: Teachers vary greatly on this. Most adhere to the academic-achievement-only criterion. Many allow apparent effort to influence grades, particularly with low achievers. Some adjust grades for ability. Teachers are naive about the importance of quality measurement. Nearly all provide equal opportunity for success by adhering to criterion-referenced assessment results.

Discrepancy: Teachers and measurement specialists differ in their understanding of the importance attached to quality measurement and in their views of the role non-achievement factors should play in grading.

3. Where do the underlying principles and actual grading procedures come from?

Recommended Practice: There should be a sound pedagogical basis for the procedures used, even though many of the decisions are value-laden. That basis should arise from high quality inservice and preservice training in assessment and grading. Grading practices should be tempered with personal experience and professional values to maximize student motivation and learning.

Case Results: The principles and practices certainly do not arise from training for any of the teachers studied. None had had formal coursework in this area and most saw this as problematic. Some practices are spelled out in school or departmental grading policy. Others arise from personal classroom experience as a teacher and as a student.

Discrepancy: Relevant, useful training in grading practices is non-existent for most teachers; technical assistance in this arena is almost never provided.

4. What factors outside the classroom influence grading?

Recommended Practice: District policy should be written to bring common meaning to grades. Those policies should be clearly communicated to teachers, students and parents.

Case Results: Grading policy often influences practice. Policy often determines frequency of reporting and the form in which information is reported. Administrators also sometimes try to influence grades, as in cases where a teacher's distribution of grades is reviewed by the principal to be sure grades are not too high or low. Parents exert influence on occasion, as in instances where a teacher's grades are seen as too low. Department norms were also mentioned as a factor, as when the department chair compares grade distributions in an investigation of consistency across department members.

Discrepancy: Policy often influences grading practice. However, so do a great many other factors unrelated to actual student achievement--the key factor.

5. How do you establish your standards of achievement expectations?

Recommended Practice: Two factors should have influence: (1) the instructional materials adopted for district use, and (2) high-quality information on prior student learning. Teachers should have realistically high achievement expectations for all students, tempered by knowledge of student entry characteristics.

Case Results: The textbook, department standards and years of personal experience allow teachers to establish expectations. Some mention that they adjust expectations based on student performance early in the year, others are not influenced by this. Some report that they are influenced by long-term trends in student performance; others hold doggedly to their standards despite changes in students.

Discrepancy: Those whose expectations are not influenced by student entry characteristics or by ongoing data on the student's rate of achievement violate principles of sound instruction.

6. Are expectations individual or the same for all students? Do they vary with achievement level?

Recommended Practice: Expectations should be a function of the student's beginning state of knowledge and skill, with students grouped according to common beginning states. Expectations and grouping should be based on quality achievement measures.

Case Results: Expectations varied for students in a given class and they vary greatly across classes at different levels of prior achievement. These variations were a function of teacher intuitions about student ability and effort.

Discrepancy: Teachers may rely on intuitions about ability and effort instead of systematic assessment of prior learning.

7. Are these expectations communicated to students?

Recommended Practice: Standards should be communicated to students.

Case Results: Grading practices are commonly communicated in writing or at least in a discussion with students; performance expectations are sometimes communicated via samples of good work.

Discrepancy: Grading practices and performance expectations are not the same thing. Some teachers are not aware of the differences. Performance expectations are often vague and not clearly understood by students.

GRADING PRACTICES

1. What student characteristics of those listed (A through F) below are considered in determining a student's grade?

A. Achievement--measured acquisition of knowledge and skill

Recommended Practice: Achievement should be the central ingredient.

Case Results: It is very important to all teachers.

Discrepancy: None.

B. Ability--measure of learning potential

Recommended Practice: This is very difficult to measure well in the classroom. Therefore, adjusting grades for ability is not recommended.

Case Results: Ability is a factor for some teachers. It tends to influence expectations and grades for low and high achievers. High achievers tend to be graded on achievement alone, low achievers on achievement and effort.

Discrepancy: Teachers may consider ability on the low and high end. However, they tend to not "measure" it themselves. They rely on student placement results (i.e., students placed in low groups by counselors) to be their index of ability. Expectations are adjusted down for low achievers and up for high achievers, rendering the meaning of any particular grade unclear.

C. Attitude--strong or weak, positive or negative feeling about the content, teacher and/or school in general

Recommended Practice: This will be reflected in the amount the student learns and therefore will be raflected in achievement; therefore, it should not be assessed again and factored into the grade. Sound practice is to assess it by itself and report results separately or ignore it.

Case Results: Teachers report striving not to consider this, but they add that it does exert subtle influences, particularly when students are on a grade borderline. It is measured via obaervation of student behavior.

Discrepancy: Teachers consider attitude in grading and measurement specialists recommend that they not do so. The message to students may be "show a positive attitude" and you don't have to learn as much.

D. Motivation and effort--amount of hardwork, initiative and perseverance invested in study

Recommended Practice: This too will exert influence on amount learned and will show up in achievement; it can be the topic of additional, separate feedback to the student, hut should not be factored separately into a grade.

Case Results: Teachers give this factor significant weight, particularly for the borderline student. They measure it in percent of homework completed, extra credit work and judgments of the "appearance of trying hard."

Discrepancy: Teachers consider this factor in grading over and above achievement despite recommendations to the contrary. The message sent to students may be "if you appear to try hard" you don't have to learn as much.

E. Interest--predisposition to like or dislike a particular subject

Recommended Practice: Again, this will be reflected in the amount of energy devoted to study and therefore in achievement; it should not be factored into the grade separately.

Case Results: Many of the teachers in this study teach required material. They sense that many students aren't there out of interest in being in class. They strive not to consider this factor.

Discrepancy: None.

F. Personality--personal emotions, behaviors and/or characteristics unique to individual students

Recommended Practice: Teachers and students, like other people, will like and dislike those around them as people. This has no role to play in grading. Achievement should speak for itself.

Case Results: Many teachers described strategies they use to prevent this from influencing grades, such as considering only achievement, rating performance in the blind, and involving another teacher in key decisions. They t[illegible] not to consider these factors in grading.

Discrepancy: None.

2. Of the alternatives listed below (A through E), what measurement methods are used in determining a grade?

A. Daily written assignments

Recommended Practice: These are only useable in grading to the extent that they are evaluated to reflect how much of the required material the student has learned. Grading simply based on activities completed is inappropriate.

Note: Some contend that homework should be an instructional device, not an assessment device. It should be practice in preparation for assessments and should not be graded.

Case Results: Teachers give daily assignments a great deal of weight in grading. They contend that this keeps students motivated to do the work. As assessments, assignments are typically very brief.

Discrepancy: Grades may reflect amount of work completed more accurately than amount learned. Also, brief assignments may suffer from a lack of validity and reliability as assessment tools, while remaining valuable learning aids. Generally, far too much emphasis is placed on homework in grading, according to measurement specialists.

B. Paper and pencil tests and quizzes

Recommended Practice: These should be central to grading when designed and constructed to reflect intended achievement outcomes with validity and reliability. This is true regardless of subject matter area.

Case Results: Teachers use them to measure acquisition of information and to motivate students to study. They are central to grading, but in many cases are no more important than daily assignments and affective factors.

Discrepancy: This varies across teachers. Discrepancies that exist relate to emphasis given to tests relative to other less desirable assessment practices used by some teachers. Tests can measure far more than the recall of information. Many teachers don't understand this.

C. Oral questions

Recommended Practice: These are only appropriate in grading if these conditions are satisfied:

1. The teacher must sample student answers to a representative sample of questions over time to avoid bias.

2. Student responses must be interpretable in terms of achievement (i.e., amount learned).

3. Systematic records of student performance must be kept.

Case Results: Most teachers use oral questions more as instructional than as assessment devices. Some use them to get a sense of group performance--not individual student status. A few include responses to such questions in grades. Those who do are not aware of the requirements listed above.

Discrepancy: Standards of quality assessment are not adhered to when oral questions are used in grading.

D. Performance assessments (including observations of student products, behaviors and interactions)

Recommended Practice: These are appropriate tools for grading when rules of sound assessment are observed:

1. Exercises sample the domain fairly.

2. Desired performance is clearly defined.

3. Performance criteria are articulated and communicated to students.

4. Objectivity is built into rating procedures.

Case Results: Examples were found of the use of this methodology in language arts (evaluation of writing samples), science (lab work) and in "participation grading." The extent of adherence to rules of sound assessment varied greatly.

Discrepancy: In general, teachers are unaware of and do not apply rules of evidence for sound performance assessment. The major weakness is the lack of clear performance criteria and the consequent inability to communicate those clearly to students and evaluate objectively in terms of them.

E. Assessment of amount of or rate of work completed

Recommended Practice: This is appropriate in grading only when it serves to reflect systematically the achievement that results from doing the work.

Raising a grade just because the student did extra credit work is not appropriate unless completion of that work provides concrete evidence of greater learning.

Case Results: Teacher consideration of this took two forms: Raising or lowering the grade based on number of assignments completed, and raising grades for completion of extra work.

Discrepancy: Sometimes teachers assume that just doing the work produces learning. Assessment specialists claim that learning outcomes must be verified via some form of evaluation.

3. How much evidence is gathered before assigning a grade?

Recommended Practice: We need to balance economy of time for assessment, quality of assesement and amount of data gathered. Enough evidence should be gathered to make the teacher confident that (s)he can estimate how much of the required material has been learned. Certainly more than one or two quality assessments are in order.

Case Results: Patterns vary from 2 to 6 major samples of student work per grading period, with numerous intermittent smaller samples including daily assignments. Most teachers seem to gather a great deal of information.

Discrepancy: Many teachers seem to believe that more data mean valid and reliable data, regardless of other qualities of assessment. This is simply not the case. Often, too much data are gathered, using up valuable instructional time.

4. What factors influence the acceptability of evidence to be included in the grade?

Recommended Practice: The key factors should be validity (match to intended outcomes), reliability (dependability) and cost (time required to gather).

Case Results: While teachers strive to match content taught in tests, they very often fail to use assessments that reflect cognitive levels of instruction (e.g., teach analysis, test recall). Issues of reliability are almost totally ignored. Assessments are designed to be completed and scored within certain time constraints.

Discrepancy: Issues related to errors of measurement and objectivity are almost never given any attention. Also one form of invalidity--discrepancies between cognitive levels of instruction and assessment--is often ignored.

5. How are various grading ingredients combined? What weighting schemes are used?

Recommended Practice: Differential variation across ingredients should be eliminated before weighting. Preset weights should be established rationally and applied in a preplanned manner.

Case Results: Teachers universally convey a weighting scheme in terms of the percent of the grade they want determined by each factor. Methods for achieving that end vary greatly. Some accumulate records, scan them visually and estimate a grade. Some compute mean scores and convert to a grade. Some accumulate points over a term (sometimes via computer) and convert. Few can verify that their intended weighting schemes are, in fact, carried out.

Discrepancy: Appropriate weighting schemes are not understood or used.

6. Of those listed below (A through D), what standards are used to interpret student performance and convert to a grade?

A. A prior assumption that grades will be distributed normally

Recommended Practice: This practice is indefensible, as achievement is not naturally distributed in such a predictable manner within classroom groups. It varies as a function of material to be learned, student investment in learning and the quality of teaching.

Case Results: Only one of the teachers studied used this system.

Discrepancy: None.

B. Fixed percent cutoff scores

Recommended Practice: This is acceptable if the scale subdivided by cutoff scores arises from valid and reliable measures, and if everyone involved understands and uses the same cutoff scores. This allows a grade to systematically and consistently reflect the proportion of material (in whatever subject) learned.

Case Results: Fixed percent cutoff scores are used by most teachers. In fact grading policy often specifies the percent cutoffs for each grade. However, underlying quality of measurement varies greatly, cutoff scores often vary greatly, and the manner in which averages (weighted or not) are computed varies greatly. In addition, the role of non-achievement factors in grades varies a great deal from teacher to teacher. Therefore, the meaning of a grade is never clear.

Discrepancy: The use of percent cutoff scores creates the appearance of systematic, objective, rigorous grading. However, a look beneath the surface reveals that this is not often the case, because cutoff scores are often fixed arbitrarily and cutoffs vary greatly even within schools.

C. Accumulation of points over time

Recommended Practice: This is acceptable if points systematically reflect amount learned--not just amount of work completed.

Case Results: This system is used by many teachers; some accumulate points on a computer.

Discrepancy: Accumulating points can have subtle weighting effects on grades that users may not be aware of. Fixing cutoff scores on point totals can be quite arbitrary, making interpretation difficult.

D. Fixing grade cutoff scores at gaps in the distribution of student performance

Recommended Practice: Such gaps represent arbitrary setting of post hoc standards and thus resulting grades are difficult to interpret.

Case Results: This method was not used among the cases studied.

Discrepancy: None.

7. What is the rationale for choosing the cutoffs used?

Recommended Practice: The standard should be chosen on the basis of a clear understanding of the appropriateness of a criterion-referenced grading system. If adopted to comply with policy, one would hope that the teacher would understand the assumptions underlying the policy.

Case Results: Teachers tend to see the percent cutoff system as equitable, straightforward, and easy to understand and use. For these reasons, they receive few complaints. It appears to be an objective system.

Discrepancy: None.

8. What procedures are used to decide the grade assigned a student who is on the borderline between two grades?

Recommended Practice: The borderline case should be clearly defined in a standard way in terms of the composite of all achievement indicators. The decision maker should recognize the fact that the composite contains measurement error and is therefore imprecise. The decision should be made on the basis of additional achievement indicators not already included in the composite, if possible. If that's not possible, performance on assessments related to the most important objectives should be reevaluated. In no case is it defensible to make borderline decisions with non-achievement information. The purpose of reviewing borderline cases, in practice, is to ensure that students will not be penalized for extenuating circumstances that may have negatively affected their achievement indicators.

Case Results: The decision is based most often on subjective factors unrelated to achievement (predominantly effort) and most often results in the student being moved up rather than down on the grade scale.

Discrepancy: Measurement specialists and teachers differ in their opinions as to the appropriate data to consider in making this decision and the purpose of reviewing borderline cases.

IMPACT OF GRADING

1. How do grades affect students:

A. Achievement (amount learned)

Desired Affect: Grades probably affect students at different levels of achievement differently. High achievers are rewarded by grades and strive to

achieve more. Low achievers may see grades as consistent evidence of their failures and may reach a sense of futility; thus achievement will suffer. This can happen in early grades. In general, grades probably enhance achievement by enhancing motivation for some, but not all students.

Case Results: Some teachers contend that grades cause students to focus only on what is tested. This can narrow the scope of what is learned. Therefore, they dislike grades and grading. However, most contend that grades enhance achievement by making students work harder to meet targeted expectations.

Discrepancy: Teachers are aware of the complex relationship between achievement and grading, particularly the effect of grades on low achieving students. However, grading mandates make it difficult for them to respond accordingly.

B. Motivation (willingness to work)

Desired Affect: Grades can and should be used to motivate students to study and learn. If underlying assessments reflect significant (not trivial) learning outcomes and do so well, the motivation to attain a high grade can result in significant learning. But if the assessment practices and underlying assumptions of grading are unsound, the effect of grades on motivation becomes very complex. Motivation may suffer in this case because the student sees no pay off for working--not because he/she hasn't learned, but because poor assessment fails to detect and reward what is learned.

Case Results: Case studies suggest that grades affect motivation, but sometimes in complex ways. For instance, the pursuit of high grades in advanced classes can motivate students to focus on the mechanics of attaining

the grade--not broaden learning. Or, students who regularly receive high grades with little effort may not be motivated to reach full potential. On the other hand, students constantly barraged with public evidence of their failure may fall further behind and give up. Thus grades reduce motivation.

The effect of grades on motivation also can be negative if poor assessments are used in grading. For instance, students who study one body of content and are unexpectedly tested on another will sense the lack of fairness and may give up studying. Case studies provided examples of all of these effects.

Discrepancy: The complexity of the grading/motivation relationship is not clearly understood by many educators, including measurement specialists.

C. Academic self-concept (perception of one's ability to learn)

Desired Affect: Grades contribute greatly to students' perceptions of their ability to learn and be successful. Therefore, they must be based upon sound, quality assessments and grading practices.

Case Results: Students can be and often are intimidated by grading processes. Many believe that if they don't attain high grades they cannot learn. Those who are intimidated may give up and not achieve their potential. This happens to many and teachers often give up on them to concentrate on high achievers, who believe they can learn. As a result, students begin to cast themselves as B, C or D students early on and regard this as irreversible.

Discrepancy: We may not completely understand the complexity of the grade/self-concept relationship for students. Low achievers may suffer as a result.

D. Locus of control (internal sense of control over one's own well being)

Desired Affect: Grading practices should give students a sense of control over the grades they receive. Greater effort should be reflected in greater achievement, which in turn yields higher grades. When these connections are made, success breeds a sense of control and more success. When they are not, the students may lose the sense of how to behave in their own best interest. Motivation to learn may suffer.

Case Results: Teachers report that students tend to assume responsibility for the grade they receive. High achievers generally see a relationship between work and rewards. Low achievers seem not to try "for some reason."

Discrepancy: Little attention is paid to low achievers in this regard. They may never have been taught the relationship between quality effort (not just hard work) and grades.

F. Attitudes toward teacher, school, and learning

Desired Affect: Grading practices should be seen as clear and appropriate and underlying assessments seen as fair. This may not be sufficient to contribute to positive attitudes, but it is certainly necessary.

Case Results: Students do not complain about or challenge grading practices and teachers therefore assume that students understand them and see them as fair. Teachers also point out that perceptions of fairness and appropriateness may change with time. For instance, college students sometimes return home with a new sense of the appropriateness of high school assessment and grading.

Discrepancy: Educators generally do not understand how much students understand about grades or how fair they think they are. If the teacher's assumption is incorrect, however, perceptions of assessment and grading could contribute significantly to a malaise about school.

2. How do grades affect teachers' perceptions of their success as teachers?

Recommended Practice: In the ideal case, grades should serve as a clear reflection of how well the teacher is doing the job. If the teacher works harder and/or smarter, more students should achieve and grades should go up. Of course, the reverse would also be true for those who do the job less well, assuming sound underlying assessments. The ultimate goal should be for every student to master all of the material and receive an A.

Case Results: In practice, there appears to be a conspiracy among educators to call this goal into question. Distributions of grades are often reviewed by the principal, at least some of whom believe grades should be normally distributed. Among teachers studied some were pressured for having too many high grades and others for having too many low grades. The former were thought to be too easy; the latter too tough.

Discrepancy: Many educators are simply not well-informed enough about grades and grading to know what their goals should be. Sound underlying assessments are not universal. Teachers differ widely in their assessment practices and in their use of non-achievement factors in grading. Thus, grades cannot be used to show that the quality of instruction is improving.

Possible Reasons for Discrepancies

While the preceding results point out some of the commonalities and discrepancies between recommended and actual practices, the case studies did not provide information on why those discrepancies may exist. The perceived cause may depend on one's point of view. For instance, teachers may view the discrepancy and claim that recommended practice is simply unrealistic, while

measurement specialists may view the same problem and suggest that it exists because teachers don't understand sound grading practices. The real cause may be a combination of the two or it might be that best practice is really a matter of opinion depending on the context.

The issue of cause is crucial if we are to bring actual and recommended practice into line with one another. For instance, if best practice is simply a matter of opinion, we need to be sure teachers are aware of all of the procedural options available to them. On the other hand, if recommended practice is unrealistic given the constraints of the classroom, then new practices must be devised that account for these constraints. Third, if the discrepancy is at least in part a matter of inadequate knowledge of sound practice on the part of the principal or teacher, then further training is in order.

Therefore, in the chart that follows, we have listed each discrepancy uncovered above and have attempted to sort out our initial speculations as to why each exists.

REASONS FOR DISCREPANCIES BETWEEN RECOMMENDED AND ACTUAL PRACTICE

Discrepancy	Recommended practice is a matter of opinion	Recommended practice fails to account for realities of the classroom	Teachers lack knowledge of sound practice
ANTECEDENTS			
2. Teachers and measurement specialists differ in importance attached to quality measurement.			Teachers are unaware of the grading problems arising from poor measurement.
Non-achievement factors influence grades.		Teachers want to reward seriousness of purpose and effort. Grades are the only reward and punishment many have to give.	Teachers are unaware of the double jeopardy resulting from evaluating achievement, which is greatly influenced by effort, and then grading effort again to factor into the grade.
3. Professional training plays no role in determining grading practices.			Teachers have no opportunities to learn and compare grading alternatives.
4. Policy and other political factors influence student grades.		Building and department administrators often exert influence on teachers to change their distribution of grades. Under such pressure, some teachers must respond.	Often the administrators are no more thoroughly trained to institute sound grading practices than are teachers.
5. Achievement expectations rarely are determined by student entry characteristics.	Teachers try to size students up at the beginning of the year as best they can and respond as needed to different student needs.	Textbooks are adopted by grade level and teachers often must cover required material.	
6. Individualization is based on intuitions about ability and student effort rather than concrete data.	Experienced teachers gain a sense of student capabilities as a function of years of experience. Intuitions are not blind guesses.	Data on student entry characteristics are often unavailable and undependable. There is not always time to assess. Guidance counselors handle the assessment and placement.	
7. Performance expectations are often vague if they are communicated to students at all.		This varies across school subjects. Math and science expectations may be more clear; language arts, social studies, arts, P.E. may be more prone to vague criteria.	Teachers are untrained in the clear definition and articulation of the performance continuums underlying their instruction.

Discrepancy	Recommended practice is a matter of opinion	Recommended practice fails to account for realities of the classroom	Teachers lack knowledge of sound practice
GRADING PRACTICES			
1. Student characteristics influencing grades:			
A. Estimates of student ability lead to adjustments in grading standards making grades a reflection of more than just amount achieved.	If the same achievement standards applied to all, those of low ability would always receive low grades, no rewards, and no success. In addition, measurement specialists remain unsure of the construct differences between ability and achievement.	Teachers need and want to provide some rewards even to low achievers; grades are the only concrete reward and punishments have to give.	Teachers may be unaware of the achievement/ability distinction and its role in the classroom.
C. Teacher perceptions of student attitudes influence grades; measurment specialists say it should not.	Attitude is related to achievement.	Teachers want to encourage a positive attitude by rewarding it and discourage a negative attitude by punishing it. Grades provide a means of doing so.	Teachers are unaware of the fact that attitude influences amount achieved; therefore, to evaluate it seperately as a grading factor doubles it's weight in the grade.
D. Teacher reward the appearance of seriousness of purpose and effort with higher grades; measurement specialists recommend against this.	Motivation and effort are related to achievement.	Teachers want to encourage effort by rewarding it. Grades provide a means of doing so.	Teachers are unaware of the fact that effort influences amount achieved; therefore, to evaluate it separately as a grading factor doubles it's weight in the grade.
E. Apparent student interest is sometimes a factor in grading; those in measurment say it should not be.	Interest is related to achievement.	Teachers want to encourage interest by rewarding it. Grades provide a means of doing so.	Teachers are unaware of the fact that interest influences amount achieved; therefore, to evaluate it separately as a grading factor doubles it's weight in the grade.
2. Measurement methods used in grading:			
A. Daily written assignments are graded; some measurement specialists contend they should be practice only.	More practice leads to higher achievement. Teachers need to encourage practice.	Teachers contend that students won't do the work (i.e., practice) if a grade is not attached.	
Grades influenced greatly by homework can reflect work completed; not amount learned.			Teachers need to be clear about the importance of achievement as the learning criterion.
B. The importance attached by teachers to performance data from sound measurment of achievement in grading is not always as high as measurement specialists would like.		For reasons related to classroom management, teachers want to reward many facets of student performance--achievement is only one. Grades are their only concrete reward and punishment available.	

Discrepancy	Recommended practice is a matter of opinion	Recommended practice fails to account for realities of the classroom	Teachers lack knowledge of sound practice
C. Accumulated mental records of student responses to oral questions during instruction influence grades for some teachers; measurement specialists say such indicators lack quality.		Time will not allow representative sampling of performance or careful recordkeeping.	Teachers are often unaware of the quality control standards required to gather sound data of this sort.
D. Performance assessments are used, but quality control standards are not met.			Teachers are untrained in sound assessment via observation and judgment.
E. Contrary to recommended practice, grades are influenced by the amount of work completed.	Achievement is related to amount of practice (amount of work completed).		Teachers need to be aware of the fact that work completed and achievement are not the same.
3. Some teachers gather too much achievement data for grading.	Effective schools recommendations say monitor achievement continuously.		Teachers need to know that the quality of data is more crucial than the amount; more data do not equal valid and reliable data.
4. Reliability (objectivity, dependability) of achievement data is rarely a relevant consideration for teachers.		Teachers do not have time to verify psychometric quality.	But they need to know simple procedures for maximizing (not estimating) reliability.
Validity problems are apparent in classroom assessment: cognitive levels of instruction and assessment do not match.			Teachers need to know how to develop and use assessments that work.
5. Strategies for combining achievement data into a term grade are not clear nor are they clearly understood.		Appropriate weighting schemes require technical data analysis that is often beyond the capabilities of teachers and is too time consuming.	Teachers are aware of the intricacies of generating a weight composite index of achievement.
6. Standards for converting performance to grades:			
B. Arbitrary percent cutoff scores vary greatly from class to class, school to school, and across districts.	No pedagogically sound rationale exists for fixing cutoff scores. Therefore, variation is to be expected.		
C. Accumulation of points over the term with point cutoff scores is similarly arbitrary and can give undue weight to some measures of achievement.	No pedagogically sound method exists for fixing cutoff points.		Teachers are unaware of the subtle influences of this record system on the importance attached to some measures.

Discrepancy	Recommended practice is a matter of opinion	Recommended practice fails to account for realities of the classroom	Teachers lack knowledge of sound practice
8. In selecting grades for borderline cases, teachers consider effective factors, while measurement specialists urge further consideration of achievement data.	Affect is related to achievement.	Teachers want to reward seriousness of purpose and effort. Trying hard and falling just short deserves recognition.	
IMPACT OF GRADING			
1. How do grades affect student:			
A. Achievement and B. Motivation The complex relationship between and among grades; motivation and achievement are not clearly understood.	Research is needed to understand how grading practices effect motivation and achievement at all levels of the achievement continuum.	Teachers are required to assign grades to all students and want to reward good work.	Teachers need training on feedback systems that might supplement or sometimes replace grades.
C. Academic Self-Concept--Grades can cause low achievers to believe they cannot learn.	The effect of grading practices on academic self-concept is not clearly understood by educators in general.		Poor assessment can lead to students misunderstanding their achievement potential. Teachers need to know how to assess in appropriate ways to prevent this.
D. Locus of Control--Little attention is paid to the effect of low grades on the low achievers' sense of control over their own well-being.	More research is needed on this topic.		Teachers need to understand and use principles of sound assessment to prevent students from giving up due to unfair assessment.
E. Attitude--If grades are seen as fair, they contribute to a positive attitude. If they are seen as unfair, they can contribute to negative attitudes about school subjects, teachers and school in general.	Students may be predisposed to see school things like grades negatively. Besides, they may not be qualified to judge fairness.		Teachers need to know, use and make public fair assessments and grading practices so as not to risk contributing to negative attitudes.
2. Grades are not a good index of how much is being learned or how effectively teachers are teaching because they include so many ingredients and reflect so many factors unrelated to achievement.	Grades are not as effective as showing changes in student achievement as are standardized tests. Grades are motivators, not standard indicators of achievement. Therefore, we should not use grades to reflect teaching success.	Principals sometimes purport to know what a desirable distribution of grades should be. This is often assumed to be a normal distribution. Sometimes teachers must respond and grade accordingly.	Principals and teachers need training in sound assessment and grading practices and the assumptions that underly those practices.

Summary and Conclusions

The purpose of this study was to explore high school grading practices with a rather powerful microscope in order to describe practices and understand their antecedents and impacts. To reach this goal, resources were invested in the indepth case study of a very few teachers. The investigation focused on 34 grading issues. For each issue, best practices as advocated by the educational measurement community were compared with actual practices of the teachers studied. Discrepancies between the two were noted in 26 of the issues, suggesting that recommended grading practices are not adhered to in classrooms.

An analysis of possible causes for the discrepancies revealed that 21 of the 26 probably have multiple causes. At least some issues related to sound grading practices may be a matter of opinion rather than a matter of sound pedagogy (17 of 26 issues). But it is also clearly the case that recommended practice may need to be reevaluated in light of the realities of the classroom (17 of 26 issues), and training in sound grading practices for teachers and principals is needed (22 of 26 issues).

The nature and importance of each discrepancy and its true cause(s) will remain uncertain until researchers expand the range and variety of classroom assessment environments (i.e., generate more generalizable results). But for now, based on limited data, we concluded that both discrepancies and causes can be understood and that action programs to reduce differences are both feasible and necessary.

References

Bail, F.T. (1983). Rewarding practice: An alternative grading system. Improving College and University Teaching, 31(3), 109-113.

Clough, E.E., Davis, P. & Sumner, R. (1984). Assessing puppils: A study of policy and practice. Berkshire, Great Britain: Nfer-Nelson.

Cullen, Jr., F.T., Cullen, J.B., Hayhow, V.L., & Plouffe, J.T. (March 1975). The effects of the use of grades as an incentive. Journal of Educational Research, 68(7), 277-279.

Ebel, R.L. (1980). Evaluation of students: Implications for effective teaching. Educational Evaluation and Policy Analysis, 2(1), 47-51.

Ebel, R.L. & Frisbie, D.A. (1986). Essentials of educational measurement (4th Ed.) Englewood Cliffs, NJ: Prentice-Hall.

Educational Research Service. (1977). Reporting pupil progress: policies, procedures, and systems. Arlington, VA: Author.

Gronlund, N.E. (1981) Measurement and evaluation in teaching (4th Ed.). New York, NY: MacMillan.

Hills, J.R. (1981). Measurement and evaluation in the classroom. Columbus, OH: Charles E. Merrill.

Mehrens, W.A. & Lehmann, I.J. (1984). Measurement and evaluation in education and psychology (4th Ed.). New York, NY: Longman.

National Education Association. (1974). Evaluation and reporting of student achievement. Washington, DC: Author.

Sadler, D.R. (1983). Evaluation and the improvement of academic learning. Journal of Higher Education, 54(1), 60-79.

Simon, S.B. (1972, May-June). Grades must go. Pennsylvania Education, 3(5), 19-21.

Simon, S.B. & Bellanca, J.A. (1976). Degrading the grading myths: Primer of alternatives to grades and marks. Washington, DC: Association for Supervision and Curriculum Development.

Stiggins, R.J. & Bridgeford, N.J. (1985). The ecology of classroom assessment. Journal of Educational Measurement, 22(4), 271-286.