

Grades: Reporting Student Performance (Summarized from Oosterhof, A. *Classroom Applications of Educational Measurement*)

Measurement experts argue that grades should reflect student achievement and nothing else. Teachers should **not** include measures of **conduct** when calculating a summary achievement grade. Should conduct be an important area to report, provide a separate score.

The desired achievement grade will accurately portray a student's degree of content understanding. This grade, however, is sometimes confounded with various teacher actions. For example, the following should **not** occur when determining summary scores for course achievement:

- treating practice tests and routine homework as summative tests;
- administering unannounced summative tests, e.g., pop quizzes;
- altering scores to account for student effort or improvement;
- reducing summative test scores for misbehavior; and
- using extra-credit assignments to alter grades.

Additional information one should consider before assigning summative achievement scores are highlighted below.

1. What Should Be Included in a Grade?

Recall the importance of reliability and validity. Any score or grade, which is to be used in an official manner, must be valid. Listed below are points to consider that may directly affect the reliability and validity of grades.

1a. Reliability of Grades

Scores that lack consistency (reliability) also must lack validity because something other than achievement is being measured when unreliable scores result. When assessing reliability of grades used in a course, one must use one of the following techniques.

(1) Internal Consistency of Grades

Such reliability is useful for measures of single, or highly related, domains or qualities. If internal consistency is calculated for a single test that contains two or more independent domains, then internal consistency should be calculated individually for each domain. Note that this presents a problem for teachers who must report a single grade. This may require that achievement in independent areas be reported by a single grade. Beside being less informative, the single grade is likely to be less reliable, and thus less valid, for courses that contain diverse and independent content. Recall also that all types of reliability are affected by use of assessment formats that inherently have lower reliability, such as essays and portfolios.

(2) Consistency in Grading Between Teachers

Ideally teachers of similar content, or of identical courses, use the same criteria for assigning course grades. When teachers vary greatly in standards, then grades reflect less student achievement and more which teacher a student had. Teachers of similar content and courses should seek similar performance objectives and grading criteria to ensure the most accurate assessment of student achievement.

(3) Improving Reliability of Grades by Basing Them on More Scores

One should use as many measures of achievement as possible to more accurately portray a given student's achievement. This argument is identical to the notion that one use as many test items as possible to measure student understanding. Using fewer items is less desirable to using more; similarly, using fewer scores is less desirable to using

more scores to determine a summative grade. The more scores used to calculate a summative grade, the less effect atypical scores will have on validity.

(4) Reliability of Grades Based on Student Improvement

Recall the discussion of test score references (norm, criterion, ability, and growth). Grades should be based solely on student achievement, which dictates that one base grades on current status, not improvement. In addition, ability- and growth-referenced interpretations of achievement are typically less reliable than norm- and criterion-referenced.

In sum, more reliable grades result when one assigns grades according to:

- current achievement status only;
- more scores; and
- single, rather than multiple, domains.

In addition, teachers of similar content should use similar objectives and scoring criteria.

1b. Validity of Grades

Grades should be based solely on student achievement; grades should reflect only students' proficiency with course objectives. Validity is usually thought of as referring to what scores represent. Note that some also argue that validity refers to how scores are viewed or interpreted. Below are various audiences for grades and how each audience may view grades.

(1) Students

Grades should reflect only students' achievement with instructional objectives. Moreover, students should be able to anticipate, with accuracy, their course grade since they will have received multiple assessments throughout the course.

(2) Parents

Grades should reflect achievement with instructional objectives. This is important to parents as it provides an indication of their child's academic progress. A single grade will be less informative to parents than multiple grades over various academic topics (math, English, etc.) and non-academic topics (behavior, participation, etc.).

(3) Other Teachers or Counselors

Accurate grades are important over time since these will be used to monitor students' progress and achievement in different areas, so grades will be the basis of remedial and placement decisions.

(4) School Administrators

Should achievement be lacking, administrators are likely to deny promotion and the like.

(5) Other Institutions

Grades indicate whether a student has potential for success, such as in college, and where students should be placed in a new school.

(6) Employers

Research shows that grades correlate lowly with on-the-job success. Grades do, however, alter individual's choices about careers. For example, students with low high school grades are unlikely to attend college, and are therefore unlikely to become doctors, lawyers, etc.

Academic grades should not be the basis for deciding whether students can participate in extracurricular activities unless one can clearly show that the extracurricular activity is directly inhibiting student achievement.

2. Advantages and Limitations of Alternative Reporting Systems

There are several reporting formats for summative assessments. The most common is percentage and letter grades. Below are advantages and limitations of each reporting format.

2a. Percentage Grades

- convenient summary of student performance
- allows for easy computation of standard scores
- misleading information—does not indicate percentage of content mastery (typically)
- misleading precision indicated
- typically indicates only a general level of performance, although detailed levels can be achieved; focus on general levels of performance creates two potential problems:
 - sufficient information about achievement is not presented to make detailed evaluations
 - letter and percentage grades are often averaged (e.g., GPA) over diverse content and courses and this is problematic because it hides or masks non-achievement in certain areas

2b. Letter Grades

- convenient summary of student performance
- possibly an ideal level of classification; neither too few nor too many classifications of performance
- typically indicates only a general level of performance, although detailed levels can be achieved; focus on general levels of performance creates two potential problems:
 - sufficient information about achievement is not presented to make detailed evaluations
 - letter and percentage grades are often averaged (e.g., GPA) over diverse content and courses and this is problematic because it hides or masks non-achievement in certain areas

2c. Pass-Fail Marks

- convenient classification, but only for limited types of achievement measures, such as dissertations
- less able to summarize accurately student achievement levels
- students tend to achieve less under this system (lack of external motivation, grades motivate many students)
- less reliability than other grading formats with more categories for classification

2d. Checklists

- better communicate student performances, strengths, and weaknesses
- allow for reporting, separately, a variety of traits (although this, too, can be the case with letter and percentage grades)
- time consuming to develop
- time consuming to evaluate, especially for others (e.g., counselors and institutions)
- require accurate, detailed, and relevant statements regarding performance objectives

2e. Written Descriptions

- very flexible and can communicate well strengths and weaknesses
- less likely to assess well all content areas or objectives

- time consuming to develop
- time consuming to evaluate, especially for others (e.g., counselors and institutions)

2f. Parent Conferences

- allows for direct communication with parents
- representative description of all content areas or objectives likely to be absent (same limitation as found with written descriptions)
- conferences can be very time consuming (planning and executing)
- permanent documentation may also be absent from these conferences

3. Establishing Grading Criteria

The following information should be used to determine course grades.

3a. Nature and Number of Assessments

Course grades should be based upon a representative sample of all instructional objectives covered in course. The representative sample may reflect assessments of any format (written tests, performance assessments, and portfolios).

As previously explained including more items on a test provides better domain sampling of objectives and student performance. Similarly, including more assessments will provide a better measure of student achievement than fewer assessments throughout the course.

When determining which assessments to include in calculating a course grade, use only scores that result from summative assessments. Self-tests and similar methods of formative assessment should not be used to determine a final summative score.

A final course grade should reflect only achievement of course objectives. Participation, conduct, etc. should not be considered unless one wishes to provide a separate score for these attributes.

3b. Determining the Weight Given Each Assessment

Several scores from various assessments may be used to calculate a final course grade. It is likely that some scores will be judged more important than others, so these scores should represent more weight on the final course grade. The primary criterion for determining appropriate weighting should be the importance of instructional objective assessed. Student effort, improvement, or time required for each assessment should not be considered in determining weight.

Note that the more weight assigned to a given instructional objective, the more important reliability of scores for that objective. Ideally one will have more than one measure (i.e., more than one score) for important objectives.

3c. Controlling the Weight Given Each Assessment

Suppose one wishes to assign a final course grade on two assessments. The first is a written test scored from 0 to 100. The second is an essay scored on a scale from 0 to 4. Both are to represent 50% of the final course grade. If a student scores 77 on the written test, and a 3.5 on the essay, what is that student's final course grade?

The first thing one must do to calculate final grades is to set all scores on **a common metric, a common scale**, such a 0 to 100. Since the written test is already on the 0 to 100 scale, it does not require any conversion. The essay, however, must be converted to the 0 to 100 scale. In this case the conversion is $100(3.5/4) = 100(.875) = 87.5$. This is simply a conversion to percentage correct.

It is important to ensure that the diversity of scores is approximately similar for all measures. For example, with the written test, scores may have ranged from 60 to 100. However, as the reader of the essay, you may have decided to use the entire range of 0 to 4, so some students will get very low scores for this measure. Note that the converted 0 to 4 range is now $100(0/4) = 0$ to $100(4/4) = 100$. Students who scored low on the essay will receive a course grade that is low. Perhaps the essay receives too much weight, or perhaps the range of the original scores on the essay was too varied and should not have been less than, say, 2.5.

To calculate the final course grade, simply multiply the converted scores by the desired weight. For example, the final course grade for the hypothetical student will be:

$$77(.5) + 87.5(.5) = 82.25.$$

3d. Establishing Performance Standards

This issue has been discussed, to some extent, previously under score-references for interpretation. Ability- and growth-referenced standards should not be used in assigning course grades. Ideally, one will use either criterion- or norm-referenced standards for determining final grades, but criterion-referenced standards are typically viewed as best for assigning grades.

Norm-referenced may be problematic since (a) standards may vary according to the sample of students enrolled (class A's ability is less than class B's), (b) it is difficult to determine the level of achievement for norm-referenced grades, and (c) basing grades on your performance relative to others may encourage a competitive atmosphere in the classroom (although this is not always bad).

With criterion-referenced grades, one avoids the problem of varying standards and grading on a curve. Further, one can better estimate, via the final grade, the degree to which a student demonstrated achievement of course instructional objectives since one may assign descriptors or adjectives for grades, such as A = outstanding, B = above expectations, C = satisfactory, etc. Ideally teachers of similar courses will assign grades based upon similar criteria of instructional objective performance.

4. Roles of Grades in Motivating and Disciplining Students

Note that grades should reflect achievement only, and this will serve as a strong external motive for some students. One should avoid dubious roles for grades, however, such as for discipline or punishment.

4a. Use of Grades to Motivate Students

Accountability is important in education, so it is likely that grades will continue to be used in education. Further, grades have clearly become a driving force for many students, even to the extent that academic achievement and learning are secondary to obtaining high grades.

4b. Use of Grades to Discipline Students

Grades should **not** be used as a form of discipline or punishment since grades should reflect only one trait—achievement. Thus, teachers should not alter grades for:

- delinquent work
- cheating, or
- incomplete work.

Self-Test

For items 1 through 12, indicate (yes or no) whether each of the following statements regarding grading is accurate.

1. Instructors of different courses who teach similar students should be find similar distributions of letter grades.
2. The best strategy for obtaining reliable course grades is to base grades on one reliable test.
3. Grades that indicate how much each student improved during the grading period are quite reliable.
4. A valid course grade will reflect each student's achievement, determination to learn, and effort.
5. Pass-fail grades are less reliable than letter (A, B, C, etc.) grades.
6. Most consumers of a student's grades assume that the grades reflect how much the student has improved.
7. Assigning a zero to missing work negatively affects the validity of course grades.
8. The primary function of a course grade is to provide feedback to the student.
9. An advantage of percentage grades is they indicate how close a student has come to mastering course content.
10. School administrators should specify the percentage of points required for a student to obtain each letter grade.
11. When checklists are used for reporting student achievement, parents are the primary audience.
12. If grades are norm-referenced, a fixed percentage of students should be assigned a failing grade.

Answers

1. Yes – because objectives, content, and assessments should be similar
2. No – multiple assessments should be employed
3. No – growth-based references are not reliable; criterion is best for classroom grading.
4. No – grades should be based upon achievement, not effort, determination, behavior, etc.
5. Yes
6. No – most assume grades reflect achievement relative to learning objectives.
7. Yes – assigning zero to incomplete/missing work is poor practice, better to find alternative assessment of proficiency
8. No – students should be able to anticipate course grades from multiple assessments that preceded the course grade
9. No – checklists and written descriptions can provide better feedback
10. No – this should be determined by teachers who develop classroom assessments
11. Yes – but also informative to students about their strengths and weaknesses relative to objectives
12. No – normed scores do not indicate achievement of objectives, they only indicate relative performance