

Students' Grading and Evaluation: A Critical Review

A. O. FALODE¹ and C. G. UDOMBOSO²

¹*Department of Petroleum Engineering*

²*Department of Statistics,*

University of Ibadan, Ibadan, Nigeria

ABSTRACT

In the last 20 years, schools have experienced many changes in academic standards, curriculum, and evaluation. Oftentimes the term evaluation frequently has a negative connotation to learners. This has adversely affected their performances and eventual final grading. Students rarely view their evaluations as opportunities for improvement even though better performance and public accountability are the principal aims of education. Instead, evaluations are seen by learners as hurdles grounded in threat. Evaluations are barriers that channel learner thinking and behavior, frequently motivated by fear of failure, with adverse consequences for those who fall short. Such learner perceptions contrast with faculty intent where evaluation is considered a tool needed to boost student competence and protect the public. Nonetheless, learners perceive the stakes to be high and so is their anxiety. Evaluation is a process to which most learners grudgingly submit

to. It is rarely a process they seek and enjoy. This paper seeks to project educational evaluation as a tool, not as a weapon, for the purpose of improvement and mastery rather than enforcement. This outlook is expected to change the psychological climate toward constructive progress instead of apprehension.

1 Introduction

Student grades and course evaluations are important descriptors of student and faculty performance. Student grades represent instructor evaluation of students and have been used pervasively for probably as long as there have been universities and colleges. Grades often cause anxiety for students, and could cause emotional discomfort. While some students view it as a limited assessment of their work on one item, others take grades to heart and see them as reflections on their personal worth and success. As difficult as grades can be for students, the grading process is also hard for teachers and lecturers - particularly new ones. Most worries are on how much to weigh different projects, assignments and exams. However, keeping some basic principles in mind could make grading an experience that enhances student learning.

2 Definitions and distinctions in evaluation

(i) *Formative and Summative Evaluation*

Formative evaluation is done to *form* or shape the subsequent performance of a learner, specifically by generating and providing feedback. It is done during an experience, and can be done by teachers as frequently as time will allow, but it should also be done formally at specified times, for instance, halfway through an experience. Summative evaluation is done at the end of a unit of time, typically at the end, and *sums* up the student's performance. Whereas formative evaluation is done primarily for the sake of the student, summative evaluation

fulfills our responsibility to society, pronouncing the student ready for the next level of training. Summative evaluation often includes a grade as well as narrative description of performance and recommendations for improvement. A grade without comment would provide only minimal guidance to a student and would not help the student improve subsequent performance. Therefore, it is recommended that a grade (label) always be accompanied by an evaluation (description in words).

(ii) *Process and product measurements; baseline measurements*

This distinction is meant to capture the difference between the curriculum that students experience (process) and their achievements (product, outcomes). The concept is often described as the process-product paradigm. Often, our research tries to document the relationships between what we *do* to students, and how they are changed by the experience. Research shows that much of what individual students actually achieve depends as much on their personal characteristics as much as on the formal curriculum.

(iii) *Dichotomous and Scalar Grading*

Dichotomous grading divides a group of students into those who pass and those who fail. Polytomous or Scalar grading recognizes a broader spectrum of student performance by providing for a series of steps for assigning grades, such as First Class, Second Class (Upper Division), Second Class (Lower Division), Third Class, Pass, or their equivalence in Polytechnics and Colleges of Education. Continuous grading would refer to a series of numbers which have small intervals, such as 88%, 87%, 86%, etc. Generally speaking, dichotomous grading fulfills our responsibility to society by determining whether a learner is competent or not. Scalar and continuous grading helps faculty and students compare performances among students, and may also help in ranking.

(iv) *Normative and fixed standards (criterion-based)*

Normative grading is *relative* and it assigns grades to students' performance by comparing them with another group, the *norm*, such as a contemporary peer group. Normative grading can be done in a mathematical way, generating a *curve*, with grade rankings based on distance above or below the mean score. Normative grading is often done less formally, with half students in the middle (for example, a grade of Second Class (Lower Division)), a quarter receiving higher grades, while another quarter receiving lower grades. In any case, the essence of normative grading is to compare students to each other. Basically, *criterion-based* grading is really fixed-standard grading, in which experts first decide what the tested domain will be and then what will be expected standards of proficiency. This approach depends upon a prior judgment of what has *content validity*. In a fixed or absolute standard system, a group of three students working with a single teacher could all receive same grades, depending on the criteria they met. In a normative system, they are competing against each other.

(v) *Compensatory and weakest-link models*

A compensatory grading system averages aspects of a student's performance using various parameters to yield a final grade. For instance, a high score on a multiple-choice final examination plus a failing clinical evaluation might calculate to a grade of Pass. A noncompensatory (*weakest link*) approach would conclude that the student is not better than his/her lowest level of competence in a core area of evaluation. For instance, an excellent examination score would not compensate for poor professionalism, or vice versa. Therefore, a student with unacceptable performance in any domain of evaluation could not receive a passing final grade. Generally speaking, clerkships must determine which aspects of performance are so important that deficiencies in any cannot be compensated for by proficiency in others.

(vi) *Descriptive and quantitative methods (subjective and Objective)*

Descriptive methods of evaluation describe a student's performance using words.

Quantitative methods try to measure performance and yield a numerical score. Most summative grades are a combination of the two methods with some consistency in weighting descriptive methods more than quantitative ones. Both assessment methods have a role in determining summative grades and one is not inherently more valuable than the other, so the terms *subjective* and *objective* – which undervalue the former – should be avoided if possible.

(vii) *Analytic and synthetic approaches; developmental approaches*

Traditional evaluation theory *analyzes*, or *breaks up* a student's performance (to analyze in Greek is to *loosen up* or *take apart*) into several components, knowledge, skills and attitudes (or, attitudes, skills, and knowledge, *ASK*). Each component can be assessed by tools appropriate for each domain. For instance, multiple choice tests might be used to assess knowledge, and standardized patients can assess history-taking skills. A *synthetic* approach *puts things together*, and asks how the student's abilities in several domains come together to achieve a level of proficiency.

(viii) *Competence and performance*

These terms have complementary meanings, but their meanings are sometimes used interchangeably, and educators should pay careful attention to how the terms are being used in a specific context. In the more common use of the terms, *competence* is what a student has the ability to do at certain times or under test conditions (in this sense, related to the etymology of the word, to strive with, or to *compete*) and *performance* is what a student does consistently on a daily basis, even when not being watched. However, these terms can also be used in exactly the reverse senses, in which *performance* refers to a display while being observed (i.e., performing for an audience), as in being *onstage*, in test conditions, and *competence* denotes all the attributes to function independently. In a developmental model, competence can be described in relation to the steps above it (intuitive expertise), and below it (proficiency). In the

synthetic model, competence is putting all the necessary characteristics and qualities together for each patient in a sustained way. The definition of competence in a profession, in this model is the ability to give to every situation that a professional might face all that properly belongs to that situation, and no more. This means that a competent person first has to make the decision about what a situation requires.

(ix) *Reliability and Validity; Feasibility; Impact*

Reliability is the consistency, replicability, stability, or reproducibility of results (in Latin, to rely on - religare - is to trust). Reliability is the amount of the observed variance that is due to the student (true score variance) rather than the test and everything else (error variance), and is usually expressed as a decimal figure between zero and 1.0. For high stakes decisions, at least 80% of the variance should be true score variance (a reliability figure of 0.8). Validity is confidence that you are measuring what you want to measure, what you value (similar in etymology to *evaluation*). Types of validity includes:

- Content validity reflects whether assessment reflects enough of the domain you want to assess, and this can be made as a judgment of experts, or by comparison with some external standard.
- Face validity judges whether the assessment method seems to experts to be appropriate for competency in question. For instance, use of a multiple-choice test to assess interpersonal skills would not have face validity.
- Construct validity means that results are consistent with reasonable theory (e.g., experts perform better than novices).
- Criterion/concurrent validity is more numerical, and determines whether the results of your assessment method agree with other appropriate measures of students' performance.

- Predictive validity refers to whether results of one assessment measure are verified by subsequent performance, and this, too, is best demonstrated with mathematical methods, such as correlations and linear regression.
- Consequential validity is the term applied to a judgment about whether the effects of an evaluation system, typically social effects, are desirable.

(x) *Curriculum and syllabus*

To some extent, what we measure and reward will determine what students learn; in other words, *assessment drives the curriculum*. The list of topics or skills that we wish students to master is the syllabus (the term, etymologically, means *list*), and the methods we use to help students master the list, collectively, are *curriculum* (that is, the *horse race* we put students through, from *currere, to run*, as in the word *current*). This distinction has implications for evaluation.

3 General principles on grading and evaluation

- (i) Grading standards should reflect course goals.
- (ii) Grading standards should be clearly stated.
- (iii) Grading should be based only on academic performance.
- (iv) When possible, use numerical rather than letter grades.

4 The mechanics of grading

- Grades should be fair and consistent as much as possible, as well as efficiently as possible. These two intentions may come into conflict -

grading quickly can mean that the grades you give are not always fair or consistent.

- Criteria should be state explicitly. For example; Does neatness count? Is posture important? How essential is good organization?
- Several papers should be read before grading starts. This is helpful in gaining a sense of the overall quality of the papers.
- Consider *norming* papers before grading. For example, in group grading, each member can pick two or three papers at random, study and agree on a grade. Having done this, you will have a sense of the criteria for each grade - which will help ensure consistency.
- Consider grading papers anonymously. Have students turn in their work with a title or cover page with their name on it. Turn that page back on all of the assignments before you begin grading so that you will not know whose work you are evaluating.
- Grade only three to five papers at a time. Our mood and our energy level inevitably affect the grades we give. To avoid boredom or getting tired, grade no more than three to five papers before taking a break. When you resume, look at the last paper to be sure you were fair and consistent as you read it.
- When you have finished, *norm* all the papers yourself. If you grade papers alone, particularly over several days, it is helpful to group the papers according to grade when you are done. Do all of the papers in the same grade range generally have the same level of quality? If not, now is the time to make adjustments.
- For tests, problem sets and short answers, consider grading in teams. When assignments and tests have multiple problems to grade, you may

save time and ensure consistency by getting together in a team and each taking two or three problems to grade on all tests. While this may be somewhat tedious, it does allow each grader to get a sense of common problems and grade more quickly by focusing one's attention.

- Do not *split* grades. Some teachers like to give two or more grades on papers and assignments. For instance, one for content, the other for organization. However, this can suggest to students that the two are not connected, when in fact they are.

5 Commenting on student work

Commenting on student work is an important part of the educational process; simply assigning a grade is not enough -comments tell the student why he or she receives a particular grade and how he or she can improve in the future.

Benefits of comments:

- Comments justify and explain the grade you have chosen to give. Students may justifiably question you when you return a paper with a C and no comments. They are less likely to question the grade when you have provided many comments and explained the grade.
- Comments let you give students feedback for continued improvement. Students can see what it is they need to do better or differently in the future to reach the standards for the course.
- Finally, comments can motivate and encourage students. When properly written, comments on student work can inspire them to continue working and improving as learners.

Following are some suggested ways to comment on student work from a constructive and educational perspective:

- (i) Comments should be balanced.
- (ii) Avoid over-marking student work. That is, avoid over-criticism.
- (iii) Gauge the tone of your comments to promote learning. Rather than writing *no!* when a student makes a mistake or is unclear, consider more constructive comments. Sometimes it may be appropriate to raise a question or allude to other ideas.
- (iv) Explain the grade. Use your comments to help students understand why you graded their work as you did.

6 Feedback on student performance

Returning students' work does not have to be simply an administrative task that takes up valuable class time – it can in fact be a worthwhile learning experience. It provides you with a chance to give the entire class feedback that will help them see and explore relationships, applications and implications of the knowledge they have been studying. Ask students what they thought of the test or assignment. You can begin by asking students to describe what they thought the most difficult areas of the assignment were and why. This gives you a sense of where they struggled and may need extra help.

Give students a sense of the class's overall performance. Students typically like to know how they did in relation to their peers. Once you have graded all the work, see how the grades are distributed. Show students this distribution and talk in general terms about where the class seemed to do well and where it struggled. Review those areas of the assignment where students struggled. If, while you grade, you keep track of those problems or ideas with which students most commonly struggled, you can take time when you return work to review some important concepts and provide the foundation for future learning.

7 Handling student complaints

Invariably, no matter how careful and consistent you are, one or more students will raise a question about the grade they have received. Most students will simply want an explanation as to why they received the grade - a clearer sense of what your standards are. They certainly deserve this. Setting out your criteria for grades early and making comments on each assignment will help cut down on the questions you receive. However, what should you do when a student approaches you and asks to talk about his or her grade?

- (i) Keep records of student performance. It is helpful to keep your own notes on the work of each student. This will help you in tracking his or her overall performance and will allow you to be prepared should any questions arise. You should hold on to this information for several years in case there are questions even after the course is over. If a student approaches you with a question, it is preferable to set up a meeting and then arrive at the meeting ready to explain your decision.
- (ii) Come with any notes you have on the student's work, with an example of a model assignment or test, and with an explanation for how the student's work did not meet that standard. Some instructors like to ask the student to submit a paragraph explaining their question and why they think the grade should be changed.
- (iii) Listen to the student. When a student wants to talk about the grade, your first task is to listen to their thoughts. Students want to be heard, and you can go a long way to diffusing a potentially tense situation by listening to their concerns before making any decisions. It is also important to keep an open mind about the matter. We all make mistakes and a student may have legitimate concerns.

- (iv) Respond to the student. Once you have listened, it is time to respond to the student. You may show how he or she did not meet the standards. In that case, it is helpful to turn the discussion to what they might do in the future, pointing to specific examples in their work that can be improved for later success. If you think you may want to change the grade, it is acceptable to tell the student you would like to think about it, or that you need to consult with your supervisor, but tell them when you will get back to them. Students will respect your honesty and your willingness to think about it.

8 Conclusion

The successful practice of science and engineering requires the effective application of technical and entrepreneurial skills. Studies continue to document significant deficiencies in these skills areas among students. Academics need to think hard about their reason for wanting to assess a student's knowledge, technical ability, self-confidence, dependability, honesty, or any other relevant characteristic. Only after identifying the purpose of the evaluation (e.g., educational diagnosis, technical proficiency, overall performance) should the programme director select a measurement tool that will produce meaningful data to inform the needed decision.

References

1. Center for Excellence in Teaching. (1999). *Teaching Nuggets*. Los Angeles: University of Southern California.
2. Center for Excellence in Teaching. (1999). *Evaluating and Grading Student Work (1)*. Los Angeles: University of Southern California.
3. Davis, Barbara Gross. (1993). *Tools for Teaching*. San Francisco: Jossey-Bass.
4. McKeachie, J. Wilbert. (1999). *Teaching Tips: Strategies, Research and Theory for College and University Teachers (10th ed.)*. Boston: Houghton Mifflin Company.

5. Race, Phil (2007): *The Lecturer's Toolkit - A Practical Guide to Assessment, Learning and Teaching*. Third Edition. Routledge (Taylor and Francis Group), London and New York.