



OFFICE OF THE DEAN, ACADEMIC PROGRAMMES

NATIONAL INSTITUTE OF TECHNOLOGY, TIRUCHIRAPPALLI

TIRUCHIRAPPALLI 620 015

Evaluation & Grading @ NITT – A Review and a Look Ahead

Dr. Ramakalyan Ayyagari, Dean (Academic)

The grades assigned by instructors at the end of every semester, rather seriously, convey the level attained by *each* student in the class. A range of decisions – by the peers, other faculty, administration, potential employers, etc. – are made using letters of the alphabet. This calls for deploying generally-accepted evaluation policies and rationale in assigning grades; else, these letters send a wrong message leading to the downfall of a whole generation. And, such policies and practices need to be carefully formulated, as well as subjected to rigorous reviews periodically, so that they serve the intended purposes.

However, there is a serious threat faced by every university – *philosophical* debates make the grading a controversial topic. And, this is because we do not, yet, have research studies on:

- What should a course grade represent?
- What should an "A" grade mean?
- What percent of the students in my class should receive a "C?"
- Should language skills be *judged* in assigning a grade to a paper?

These "should" questions require value judgments rather than an interpretation of research data; and, we do not gain from the heated arguments by a section of faculty in a limited amount of time in a closed meeting hall, minutes before the results are published. The answer to each of these questions is subject to the instructor who taught the course during that semester; of course, year to year we see variance in the receptivity of the student groups too. Since every instructor is, rather, bound to ask similar questions and find acceptable answers to them in establishing their own grading policies, clearly, it is not sufficient to have "some" arbitrary method of assigning grades. Instead, the method used must be defensible by the user (instructor, student, parent, or any other stakeholder) in terms of his or her beliefs about the goals of University education, and not discounting the realities of the setting in which grades are given.

Quite often we see polarized views on the role of university education by the instructors. One group believes that the graduate should be among the "prestigious" group which has survived four or more years of culling and sorting. On the other end of the spectrum there exists another group which believes that students in their early adulthood should be able to earn a degree over a period of time; our regulations permit up to 6 years! Each group has different grading policies.

Behind such extremes as above, an instructor's faith in the system is influenced by many factors, and a change in the faith may have a reflection on the evaluation policy. Beyond some depth this issue might turn out to be tricky, and hence this issue is grazed on the surface in this document. First, the type of instructional strategy that has been set in a given institute over several years, in a big sense, dictates the grading procedures; a visible example – a 'mastery learning' approach to teaching is incongruent with a grading approach based on competition for an arbitrarily set number of "A" or "B" grades. Grading policies of the institute and the departments (through the PAC) are quite likely to limit the procedures conceived by the instructors, and might actually force a basic grading plan. Secondly, the recent response to grade inflation has caused some faculty, individually and collectively, to alter their philosophies and procedures. Pressure from colleagues to give lower or higher grades often causes some faculty members to operate in conflict with their own views. Lastly, it is also

possible that student grade expectations and the need for positive student evaluations of instruction probably both contribute to the shaping or altering of the grading philosophies of some faculty, particularly those who took to the profession recently.

With careful thought and periodic review, most instructors can develop satisfactory, defensible grading policies and procedures. To this end, several of the key issues associated with grading are identified in the sections which follow. It may be noted that this document is a framework, and does not impose any rule as such. In fact, this document aims to minimize the dissonance created by institutional restraints which probably contribute to the wide-spread feeling that end-of-course grading is one of the least pleasant tasks facing an instructor.

What typically prompts Grading

Across the globe, when it comes to the last leg of assigning letter grades, interestingly, some kind of comparison is made by the instructors. It could be a comparison of a student's performance to that of his or her classmates, or to the standards of excellence (i.e., pre-determined objectives, professional standards), or to a combination of parameters like these. The following comparisons are popular:

Comparisons with Other Students – Not Necessarily the Relative Grading

The system is a common one that many faculty members are familiar with. Given additional information about the students, instructor, or college department, grades from the system can be interpreted easily.

By comparing a student's overall course performance with that of some relevant group of students, the instructor assigns a grade to show the student's level of achievement or standing within that group. An "A" might not represent excellence in attainment of knowledge and skill if the reference group as a whole is somewhat inept. All students enrolled in a course during a given semester or all students enrolled in a course since its inception are examples of possible comparison groups. The nature of the reference group used is the key to interpreting grades based on comparisons with other students.

On the flip side, these grades are difficult to interpret by a third party without additional information about the overall quality of the group. Hence, there is usually a need to develop course "norms" which account for more than a single class performance.

Comparisons with Established Standards – Absolute Grading, in some sense

Everyone is all-too-familiar with the merits of this method. Most students, if they work hard enough and receive adequate instruction, can obtain high grades. The focus is on achieving course goals, not on competing for a grade. Students do not jeopardize their own grade if they help another student with course work. There are no quotas in each grade category. It is possible in a given class that all students could receive an "A" or a "B." However, this calls for a clear-cut definition of the course goals and standards, and these must be communicated to the students prior to the commencement of the course. Nevertheless, a complete interpretation of the meaning of a course grade cannot be made unless the major course goals are also available.

Comparisons Based on Learning Relative to Improvement and Ability

In some of the western universities these comparisons are found to be used by instructors in grading students. Unfortunately, they are highly debatable as they pose serious philosophical and methodological problems. Broadly, comparing relative to improvement is looking at how a student (or a small group of students) have improved from the commencement to the end of the course. It is possible that someone performing consistently throughout gets penalized while one (perhaps deliberately) showing poor performance early on the course and improvement towards the end gets rewarded. The other comparison on ability is about how much a student is expected to learn and how much has been measured (by way of assessments) during the course. In the early days of AIEEE (perhaps even now), NITT faced the dilemma – best among the country come to our institute how would an instructor de-grade them and (statistically!) de-mean them?

It is now easy to argue that a single course grade should represent only one of these several grading comparisons, else it leads to too much of a communication burden.

Basic Grading Guidelines – The Role of chairperson of the PAC

1. Students believe that fair and explicit grading policies are an important aspect of quality instruction.
 - a. Grading plans should be communicated to the class at the beginning of each semester.
 - b. And, these grading plans should not be changed without thoughtful consideration and a complete explanation to the students.
2. Grading components should yield accurate information.
 - a. The number of components or elements used to assign course grades should be large enough to enhance high accuracy in grading.
 - b. One can do simple hand-calculation – Consider between two assessments @ 25% and 75% and five assessments at @ 10%, 20%, 20%, 25% and 25%; in the former normalization errors creep in and damage grading more than in the later.
3. Grades should conform to the practice in the Department and the Institute.

A Discouraging Note on The Distribution Gap Method

This widely-used method of assigning test or course grades is based on the relative 'ranking' of students in the form of a frequency distribution of student exam marks. The frequency distribution is carefully scrutinized for gaps, i.e., several consecutive marks which have zero frequency. A major fallacy with this technique is the dependence on "chance" to form the gaps. It is highly likely that these gaps are random owing to measurement errors (due to guessing, or poorly written answers, etc.) which dictate where gaps will or will not appear. If marks from an equivalent test could be obtained from the same group, the gaps would likely appear in different places; some students would get higher grades, some would get lower grades, and many grades would remain unchanged. Unless the instructor has additional achievement data to re-evaluate borderline cases, many students would have their fate determined more by chance than performance.

Henceforth our institute advocates a more rigorous relative grading system described as follows.

Relative Grading

A practical rule of thumb for applying statistical methods, particularly classifying a set of about 30 or more numbers into handful of letter grades, is to make use of group comparisons. Relative grading is performed on such data sets using the following generally sound procedure:

(In what follows we assume, for illustrative purpose, there are 5 assessments for a given course, @ 10%, 20%, 20%, 25% and 25% weightage)

1. **Normalizing and Adding:** The usual practice is to set the quiz/question paper to the preset weightage. However, at the end of semester, the marks obtained need to be scaled appropriately – for example if a student x scores x_1 to x_5 in the said 5 assessments, the composite marks out of 100 is *not* simply the sum $\sum x_i$, but it is $2x_1 + x_2 + x_3 + 0.8x_4 + 0.8x_5$, i.e., $(\sum(x_i/w_i))/N$. In some resources one may find each of the components being pulled-up to its maximum, and then added; the result is same.
 - a. For a quick mental calculation assuming a student got 20 out of 25 (80%) in first assessment, and 69 out of 75 (92%) in the second. Out of 100 he should have got 86 $((80+92)/2)$, and not 89. Thus, in 5 assessments as mentioned above if the student scored 6, 16, 15, 19, and 21 his true percentage is 75, and if the instructor overlooks the weightage, it would be 77%. A counter-intuitive message to the student is that he should not neglect components with lesser percent weightage.
2. Following the regulations on passing criteria, update the data to consider only those students who are eligible for a pass.

3. Build a frequency distribution of the total marks by listing all obtainable marks and the number of students receiving each.
4. Calculate the mean, median, and standard deviation. In what follows, if the mean and median are similar in value, use the mean for further computations, else use the median.
5. The class average μ serves as the lower cut-off for the range of B's and upper cut-off for C's. Add one standard deviation to the average, to find the upper cut-off for B's, as well as lower cut-off for A's, i.e., $\mu + \sigma$. One σ above this would fix the border between A and S. Likewise, subtract standard deviations from the mean successively to find the ranges of C's, D's and E's.

Grade	Range		
S	$\mu + 2c \cdot \sigma$	to	max.
A	$\mu + c \cdot \sigma$	to	$\mu + 2c \cdot \sigma$
B	μ	to	$\mu + c \cdot \sigma$
C	$\mu - c \cdot \sigma$	to	μ
D	$\mu - 2c \cdot \sigma$	to	$\mu - c \cdot \sigma$
E	$\mu - 3c \cdot \sigma$	to	$\mu - 2c \cdot \sigma$

- a. S grade is to be awarded as per regulations, e.g., $\geq 90\%$
 - b. Instructors will need to decide logically on the scale factor c (in the interval $(0, 1]$) values to be used for finding grade cutoffs – one, or one-half, or three-fourths of a standard deviation, for example.
6. Be aware that measurement error exists in composite marks too! Hence, prudently use the number of assignments completed or quality of assignments or other relative achievement data available to re-evaluate borderline cases.

Relative grading methods like the one outlined above are not free from limitations; subjectivity enters into several aspects of the process. But a systematic approach similar to this one, and one which is thoroughly described during the first Class Committee Meeting (instructed to be conducted within the first 2 weeks of commencing the semester), is not likely to be subject to charges of capricious grading and miscommunication between student and instructor.

Here are a few suggestions I have, based on my experience both as a student and a teacher.

- a. When learning objectives and actual learning outcomes are mismatched, adjust learning objectives and re-calculate, rather than adjusting final grades. In fact, it makes more sense to adjust learning objectives over time as knowledge of students' abilities becomes more familiar.
- b. Adjust scales to fit each assessment, rather than adjusting assessments to fit scales.
- c. As you teach a course multiple times, base the distribution on multi-year or multi-course grades, rather than on single class. Weigh tasks according to their importance in demonstrating course objectives.
- d. Attempt to establish the department-standard curve – you can play around and shift the range of marks, say by M , to arrive at a common average, and notice that the individual standard deviations are not disturbed and hence your grades.

- e. A document by name Faculty Resource on Grading on the website of Univ. of Washington, some time ago, had this important line – At the level of higher education, no student should be graded on behaviour rather than their actual achievement of course objectives. If the course objectives are to improve certain aspects of behaviour then those aspects to be graded should be explicitly stated and graded according to the student's ability to achieve those objectives.

Lastly, we look at another major issue.

Grading across Multiple-Sections

Currently, in our institute we have courses (with identical syllabus) offered to two sections with two *independent* instructors. Evidently, there are some, rather unique, grading problems associated with such courses. And, virtually everyone is often concerned about the potential lack of equity in grading standards and practices across the sections.

To promote fairness and equality, the following conditions might be established as part of course planning and monitored throughout the semester during the periodic class committee meetings (CCMs):

1. It is strongly recommended that the section instructors independently carry out the evaluation and grading; even if the same instructor handles both the sections for the course, he/she is advised to grade the classes independently.
2. The number and type of grading components (e.g., mid-term tests, quizzes, mini projects, end-semester exams) should be the same for each section.
3. All grading components should be identical or nearly equivalent in terms of content measured and level of difficulty.
4. Section instructors should agree on the grading standards to be used (e.g., weights to be used with each component in formulating semester total marks; and the level of difficulty of test questions to be used).
5. Evaluation procedures should be consistent across sections (e.g., method of assigning marks to test questions, mini projects and end-semester exams).
6. Though all of these conditions can be addressed in the course planning stage, their implementation may be a more difficult task. Successful implementation requires a spirit of compromise between section instructors and the PAC chairperson and the HoD.
7. Frequent review of instructor practices in the class committee meetings and constructive feedback to the instructors are also needed. The following guidelines contain suggestions for promoting equity in grading across multiple sections:
 - a. To establish common grading components in each course section, all section instructors should agree at the beginning of the course on the number and kind of components to be used. Agreement should also be reached on the component weighting scheme and final requirements for such courses.
 - b. To encourage instructional adequacy across sections, the instructors may come together and contribute to the construction of common tests, quizzes, or projects. This will serve to "standardize" section instruction, also.
 - c. Prior to the administration of an exam, quiz or project, all instructors should agree on established letter grade cutoff marks. This consensus helps to standardize the administration of grading procedures by reducing the number of "lone wolves" who wish not to conform to someone else's standards.
 - d. In cases where the grading of particular components is more subjective than objective (e.g., more influenced by personal judgment), organized group practice helps to unify the application of evaluation procedures. For example, instructors may wish to practice grading a stack of essay exams by circulating and discussing their individual ratings. Through such group practice the instructors

involved can compare their evaluation practices with one another and become more uniform over time.

8. Any grading or evaluation changes made in a particular section should be implemented in all sections.
9. Instructors can compare their grade distributions with the grade distributions for similar courses in the same department.

Concluding Comments

No evaluation efforts can be expected to be perfectly accurate, but there is merit in striving to assign course grades that most accurately reflect the level of competence of each student. Carefully written tests and/or assignments (homework papers, mini projects) are keys to accurate grading. Because it is not customary at the institute level to accumulate many grades per student, each grade carries great weight and should be as accurate as possible. Poorly planned tests and assignments increase the likelihood that grades will be based primarily on factors of chance. It may be argued that over the course of a college education, students will receive an equal number of higher-grades-than-merited and lower-grades-than-merited. Consequently, final GPA's will be relatively correct. However, in view of the many ways course grades are used, particularly after the student leaves the institute, each course grade is often significant in itself to the student and others.

Sources Adopted:

1. <https://citl.illinois.edu/citl-101/measurement-evaluation/exam-scoring/assigning-course-grades>
2. <https://www.astronomy.ohio-state.edu/pogge.1/Ast162/Quizzes/grade.html>
3. Stephen M Stigler, *The Seven Pillars of Statistical Wisdom*, Harvard Univ. Press, 2016.