

Deliberate Grading

Academic Affairs Committee

Charge AA2 Sub-Committee: Matthew Fluet, P. Venkataraman

2016/17 AY

A grade is an inadequate report of an inaccurate judgment by a biased and variable judge of the extent to which a student had attained an undefined level of mastery of an unknown proportion of an indefinite material.

Paul Dressel, *Grades: One More Tilt at the Windmill*, 1976

1 Introduction

Final course grades are a quantitative and permanent record of a student's performance in a course. Individual instructors assign final course grades, but consumers of final course grades include students, academic advisors, other instructors (both at RIT and other institutions), admissions committees, and potential employers. Moreover, an individual student's transcript accumulates final course grades from a variety of instructors, departments, and colleges during her time at RIT. It is important that final course grades provide consistent interpretations of student performance, facilitating both judgment about a student's future performance and comparison of students' relative performance.

This white paper is a response to faculty concerns about inconsistency in the assignment of final course grades. It seeks to correct misconceptions about final course grades and to encourage practices that mitigate potential inconsistencies.

Educators and researchers with greater expertise have written significantly about grading. This is only a brief document highlighting some of the many dimensions that an instructor should consider with regard to determining and assigning final course grades. Faculty are strongly encouraged to explore the **References and Resources**, which, in turn, include pointers to significantly more literature.

Ultimately, final (and intermediate) course grades are the responsibility of the course instructor. Recognizing that RIT is a collection of diverse colleges, departments, and programs, with many

different evaluation criteria and priorities, there can be no single system for consistent grading that is suitable for all courses. Rather, faculty should be deliberate in their grading practices: familiarize themselves with the variety of assessment, evaluation, and grading strategies available, especially those of colleagues who have or are teaching the same or related courses; clearly document grading policies in course outlines and course syllabi; regularly review and compare grade distributions of multi-section, repeated, and related courses; discuss grading practices at the department and college levels. Deans and department heads are especially encouraged to note any trends or anomalies in final course grade and bring them to the attention of the appropriate faculty.

2 Background

Final course grades are governed by Policy D05.0 (Grades), particularly Section II.A:

Grade	Description	Quality Points
A	Excellent	4.00
A-		3.67
B+		3.33
B	Above Average	3.00
B-		2.67
C+		2.33
C	Satisfactory	2.00
C-		1.67
D	Minimum Passing Grade	1.00
F	Failure	0.00

An “F” grade does not count toward residency requirements (see Policy D12.0 (Graduation Requirements)) at the undergraduate level.

C- grades and below do not count toward the fulfillment of program requirements for a graduate degree.

Furthermore, Section I states that it is the instructor’s responsibility to “Define criteria for evaluation” and to “State the process for converting the professor’s evaluation criteria to the RIT grading system.”

There is no other explicit meaning assigned to final course grades by RIT policies. There are policies that convey implicit meanings through the interpretation of term, cumulative, and

program GPA. For example, Policy D05.1 (Academic Actions and Recognitions) establishes the following minimum and maximum GPAs for various interventions and honors:

- term and/or cumulative GPA ≤ 2.00 : academic probation or suspension for undergraduate students
- cumulative and/or program GPA ≤ 3.00 : academic probation or suspension for graduate students
- term GPA ≥ 3.40 (and no grades of D, F, or Incomplete): Dean's List for undergraduate students
- cumulative GPA ≥ 3.40 : graduation with honors *cum laude* for undergraduate students
- cumulative GPA ≥ 3.60 : graduation with honors *magna cum laude* for undergraduate students
- cumulative GPA ≥ 3.80 : graduation with honors *summa cum laude* for undergraduate students
- cumulative GPA ≥ 3.85 : eligible for Outstanding Undergraduate Scholar award

Also see Policy D12.0 (Graduation Requirements) and the requirements of individual programs. Finally, implicit meaning is conveyed through the use of course prerequisites, which may require a minimum final course grade in a prior course for enrollment.

As most faculty are aware, starting with the 2014/15 Academic Year, RIT has used the above Refined Grading System (RGS) (also known as the "Plus/Minus Grading System") for final course grades. The former grading system (also known as the "Whole Letter Grading System") offered only the grades **A**, **B**, **C**, **D**, and **F** (with the same quality points as above). The decision as to whether and how to exploit the increased flexibility in final course grades afforded by the Refined Grading System is left to each individual instructor; the transition was not accompanied by specific guidance on the intended usage of plus/minus letter grades in comparison with whole letter grades. Note, however, that the former grading system also had no specific guidance on the intended usage of whole letter grades beyond that found in Policy D05.0 (Grades).

During the 2015/16 Academic Year, the Academic Affairs Committee was charged to examine possible changes to the RIT grading system, including further refinement of the **A** and **D** grades. Based on submitted final course grades, it appears that over 90% of instructors are making use of plus/minus letter grades. The committee sent a survey to current instructors to gauge their practices, needs, and concerns with respect to the current grading system; the survey included both quantitative (e.g., multiple-choice) and qualitative (e.g., short-answer) questions. Based on the results from 462 respondents (276 tenure-track faculty; 114 lecturers; 64 adjuncts; 8 did not self-identify rank), the committee concluded that no changes to the grading system were necessary. There was no majority support for **A+**, **D+**, or **D-** grades and a non-trivial number of text responses that objected to further changes to the grading system.

Responses to the qualitative short-answer questions indicate that there are various opinions and practices regarding the plus/minus grading system. Some instructors felt that there is a lack of consistency in the use of plus/minus letter grades across the institute. A variety of mappings

Grade (Legacy - 2138 and prior)” organization-level scheme. Scheme 2 and Scheme 3 are also sufficiently common that they are automatically provided as the “Letter Grade Template” and the “Letter Grade - COLA” organization-level schemes, respectively. Scheme 4 is not currently automatically provided as an organization-level scheme, but it can easily be defined for a course shell. However, the existence of organization-level schemes should not be taken as institute endorsement of those schemes over other schemes; they are simply provided as a convenience and the name “Letter Grade Template” emphasizes that it is meant to be adapted, not adopted without careful consideration. An instructor always has the freedom to establish grading policies and schemes that are best suited for individual courses.

Another misconception is that there is a predictable effect on student GPA due to the transition from whole letter grades to plus/minus letter grades. A typical argument starts from the assumption that all instructors assigned final grades under the former “Whole Letter” grading system using Scheme 1 and that all instructors assign final grades under the current “Plus/Minus” grading system using Scheme 2; since a student previously obtaining a **B** for 3.00 GPA would now obtain either **A-** for 3.67 GPA, **B+** for 3.33 GPA, or **B** for 3.00 GPA the net effect would be an increase (or, at least, no decrease) in GPA. The principal fallacy in this argument is the assumption that all instructors use (or should use) Scheme 2 under the current “Plus/Minus” grading system; other fallacies include the assumption that all instructors used Scheme 1 under the former “Whole Letter” grading system and the (often implicit) assumption that students’ numeric grades are randomly and uniformly distributed. There is also often an underlying opinion about the desirability of increasing or decreasing student GPA. Finally, the hypothetical GPA of a student under the former “Whole Letter” grading system is of little relevance to those students who have matriculated under the new “Plus/Minus” grading system.

Two related misconceptions are that it is “harder” to achieve a 4.0 with plus/minus letter grades than with whole letter grades and that it is “unfair” to have an **A-** that cannot be “balanced out” by a **A+**. Again, these misconceptions arise from assumptions that common grade schemes are used by all instructors under the two grading systems and an underlying opinion about the desirability of awarding the highest GPA.

A benefit of the current “Plus/Minus” grading system is that gives an instructor the option, but not the requirement, to make finer distinctions in the performance of students. While none of the schemes presented above can be put forward as the “one-size-fits-all” grading scheme, each of them is a justifiable adaptation of Scheme 1 from the “Whole Letter” grading system to the “Plus/Minus” grading system, based on whether one considers Scheme 1 to anchor the earned GPA to the whole, low end, middle, or high end of the numeric range. Furthermore, there is no requirement and no assumption that instructors use a 0-100 numeric scale to assign final course grades.

4 About Grades and Grading Practices

Walvoord and Anderson (1998) “urge faculty to abandon false hopes that grading can be easy, uncomplicated, uncontested, or one-dimensional.” Assigning a final course grade is only the last step in a complex process of assessment and evaluation that is occurring during the entire term. It is to be expected that an instructor’s ideas about grading will evolve during her career in response to individual experiences. (Pollio and Humphreys (1988) note that college professors tend to have received better grades than their student peers, which cautions instructors against assuming that their own students view grades in the same manner that they did.) Nonetheless, all instructors should be aware of some of the expert advice and research about grading and use it to inform their own opinions. Again, faculty are strongly encouraged to explore the **References and Resources**, which, in turn, include pointers to significantly more literature.

Svinicki and McKeachie (2013) emphasize that grades are a method of communication and must capture both a historical aspect (the student’s past performance in that course) and a predictive aspect (the student’s future performance in new situations). They also highlight that different audiences use grades for slightly different purposes: Students use grades to make decisions about expected grades and performance in subsequent courses and about potential success in a career. Academic advisors and admissions committees use grades to make decisions about a student’s preparation (motivation, knowledge, skills and ability) for advanced courses and advanced study. Pollio and Humphreys (1988) remind the reader that grades did not always have the five major categories **A** through **F** and that grades were not always as pervasive a part of academic life as they are now.

The ability of final course grades to accurately predict future success is somewhat debated. Rotenberg (2010) and Pollio and Humphreys (1988) cite a number of studies that suggest only a very small positive relationship between grades and achievement; furthermore, college instructors place significantly more confidence in grades as a useful and lasting measure of absolute and relative performance (i.e., the differences between an **A** and a **C** student would persist five years or more) than do business recruiters. Svinicki and McKeachie (2013) counter that many studies look at situations where decisions were made using both grades and other predictors and low correlation is to be expected between each selection variable and the outcome.

Students’ views on grades are sometimes different from instructors’ views. Walvoord and Anderson (1998) identify four legitimate roles that grades play: “evaluation,” “communication,” “motivation,” and “organization;” they also identify three additional roles that students often assign to grades: “reward for effort,” “ticket to upward mobility,” and “a purchased item that has been paid for.” These additional roles often arise when a student wants a grade changed; Davis (2009), Rotenberg (2010), Svinicki and McKeachie (2013), and Walvoord and Anderson (1998) each devote chapters or sections to discussing grades with students.

It is undeniable that the student audience for grades is different from other audiences; not only is a grade communicated to a student, it is a communication about the student’s performance.

Moreover, while other instructors, admissions committees, and potential employers only see a student's final course grade, the student receives a multitude of intermediate course grades for individual assignments/exams in addition to the final course grade. It goes without saying that grading policies in course syllabi should clearly state how grades on individual assignments/exams contribute to the final course grade. (A few methods for such aggregation are described below.) A balance must be struck between minimizing emphasis on grades and clear communication about grades.

Experts caution against assuming that grades are the sole, or even significant, motivation for students. Svinicki and McKeachie (2013) point to literature suggesting that moderate grade motivation and high intrinsic motivation are correlated with achievement (more so than high grade motivation alone) and that students are most motivated when success can be achieved with reasonable effort. However, it is clear that intermediate course grades can serve a formative role, as one of many aspects of instructor feedback that can be used to guide further improvement in the course. On the other hand, final course grades are necessarily summative.

Svinicki and McKeachie (2013) note that grades, as measurements, should be both valid and reliable. A valid measurement is one that actually measures what is intended (or claimed) to be measured. A reliable measurement is one that is consistent across situations that do (or should) not effect what is being measured. These qualities are perhaps most easily evaluated for intermediate course grades for individual assignments/exams, but are equally important for final course grades. Certainly, a final course grade is no more valid or reliable than the validity and reliability of the intermediate course grades that are aggregated, but care must be taken to preserve these qualities during aggregation. For example, class attendance is a valid (it does, in fact, measure a student's attendance) and reliable (it would not change if measured by a different instructor) measure, but a final course grade based primarily on class attendance would not be valid (it does not measure the intended measurement communicated by a final course grade: the student's performance with respect to the course content). There is a vast literature on designing effective assignments and exams to assess student learning (again, Davis (2009), Rotenberg (2010), Svinicki and McKeachie (2013), and Walvoord and Anderson (1998) each devote chapters or sections to assessment through assignments and exams, along with further useful references).

As Walvoord and Anderson (1998) note, "your model for weighting various components is also a communication to your students about what you think is most important and about where you want them to put their effort." As such, there is such an individualized aspect to the weighting of intermediate course grades that the literature offers little universal advice. Svinicki and McKeachie (2013), indirectly, suggests working backwards: start with the goals and student learning objectives of the course; determine appropriate assessments for each objective; strive for variety and balance in assessment methods and coverage of content areas. A final step of weighting the individual assignments and exams seems to be implied. There are pros and cons to giving significantly more weight to work at the end of the course than at the beginning.

Somewhat more advice (though little research) is given on methods of calculating and assigning course grades. Walvoord and Anderson (1998) offer three distinct models:

- **Weighted Letter Grades:** where average letter grades in categories are themselves combined in a weighted average for the final course grade. This model purposefully minimizes variance of performance in a category (e.g., a “high” C average on tests and a “low” C average on tests contribute equally to the final course grade). High performance in one category cannot typically offset low performance in another category.
- **Accumulated Points:** where (continuous) numerical “points” are assigned to categories and the final course grade is determined by ranges of accumulated points. High performance in one category can offset low performance in another category. In some variations, the total points available can exceed that required for an A, thereby giving students some choice in assignments and exams.
- **Definitional System:** where the final course grade is determined by meeting or exceeding a standard for each category of work. High performance in one category cannot offset low performance in another category.

Numerous variations on these models exist: incorporating penalties and/or extra credit, establishing ceilings and/or floors, dropping lowest assignments/exams in a category.

Experts spend considerable text comparing and contrasting criterion-referenced grading (“grading against an absolute standard”) and norm-referenced grading (“grading on a curve”). The models above are examples of criterion-referenced grading, where the final course grade earned by a student is independent of the final course grade earned by other students. The typical model of norm-referenced grading is one where a fixed percentage of students receive each of the letter grades (e.g., 7% receive **A**, 24% receive **B**, 38% receive **C**, 24% receive **D**, and 7% receive **F**), but there are variations.

After a lengthy discussion, experts nearly universally argue against norm-referenced grading: grades are not a limited commodity; student learning and performance (the measurement to be communicated by a final course grade) is not a random variable following a statistical distribution; course rosters are not a random sample of the student population; and, the practice induces competition, rather cooperation, among students. The arguments in favor of norm-referenced grading are rather more opaque; the suggestion is that (explicit or implicit) administrative expectations about grade distributions pressure faculty to meet those expectations (or, at least, be required to explain in detail deviations from those expectations). Thankfully, such administrative expectations are rare to absent at RIT, though there is occasional scrutiny of **DFW** rates for introductory and service courses (precisely those courses where consistency and final course grades as prerequisites for advanced courses are most important).

While the prohibition against norm-reference grading is understandable, there are nuances. Blind adherence to a criterion-referenced grading model presupposes that the assessment methods employed are known to be valid and reliable. This might be achieved for the “perfected” course, where content and assignments/exams do not change from term to term, but does not

satisfactorily account for the natural (and necessary) evolution of a course. Most instructors will have had the experience of writing a new assignment or exam that was much more difficult than intended. How did the instructor recognize that the assignment or exam was much more difficult than intended? One way is by observing that the grade distribution for this assessment was significantly different from those of the assessment that it replaces in previous instances of the course, possibly taking into account other evidence of the relative performance of this class to those of previous instances. The instructor might respond to the situation by norming (relative to previous instances of the course) the scores for this assessment, while remaining faithful to the criterion-based grading policy set out for the course as a whole.

Davis (2009) describes a hybrid criterion- and norm-referenced grading model, where the average score of the students in the top 10% of the course is calculated and final grades are assigned based on an individual student's score relative to that average score. Hanna and Cashin (1988) give a balanced perspective on criterion- and norm-referenced grading, suggesting that grade distributions, when aggregated over many sections and terms, can inform norming of a course instance when paired with well-chosen anchoring assessments.

Related to grade distributions is the issue of grade inflation. Rotenberg (2010) notes that there have been no definitive studies on the issue. He also points out that, as the number of available majors increases, students are able to complete their degrees with courses that match their interests and motivations, naturally leading to higher grades without a lowering of standards. Walvoord and Anderson (1998) cautions that grade inflation is a national problem that should be addressed on a national level.

None of the preceding comments should be taken to imply that the grade distribution of a course is without meaning. Simply that the grade distribution should be viewed as an output of the complex interactions of grading policies, course designs, assessment methods and evaluation standards, class rosters, and other elements; it should not, without careful consideration, be viewed as an input. It is important that instructors regularly review and compare grade distributions as a means of assessment of courses and curricula.

5 Conclusions and Suggestions

The closing remarks of Pollio and Humphreys (1988) are worth quoting in full:

Sustained and thoughtful faculty discussion of grading in relation to testing and course requirements is important – but not to bring uniformity to our practices or to coerce colleagues into procedures antithetical to their values. Rather, attention to these issues brings greater clarity to classroom values and procedures; suggests new approaches to grading, teaching, learning, and testing; and promotes a greater collegial understanding of these matters as students experience them in specific individual classes.

Ultimately, final course grades are the sole responsibility of the course instructor. A single, institute-wide system for consistent grading would not be appropriate for the diverse collection of colleges, departments, and programs found at RIT nor would a top-down imposition of grading practices be effective. Rather, consistency is achieved bottom-up through the activities of peer instructors coordinated at the department- and college-levels. Moreover, if instructors are deliberate in their decisions about grading, then differences should not be taken as inconsistencies, but rather as part of the ongoing self-dialogue that characterizes higher-education.

We conclude with some, perhaps obvious, but worthwhile thoughts and suggestions to faculty and administrators:

- An instructor should reflect on the variety of grading systems and grading schemes available and make a deliberate choice of grading policies that are consistent with her teaching philosophies, recognizing that colleagues may come to different conclusions.
- An instructor should think about final course grades at the start of each course. The end of the term is a stressful time for faculty and students, made more so by fretting over assigning final course grades.
- An instructor might (informally) extend the intended course-level student learning outcomes and associated assessment methods (found in the official course outline) with rubrics that describe what is expected at various grade levels. In addition, an instructor should note any student learning outcomes that carry significantly more or less weight than the others, such that an overall final course grade may not imply the same level of mastery for a particular outcome.
- An instructor should not feel obligated to make use of all of the final course grades allowed by Policy D05.0 (Grades). An instructor's chosen grading system and grading scheme should not make finer distinctions than those measured by the assessment and evaluation tools used in the course.
- An instructor should endeavor to communicate clear grading policies and evaluation criteria to students (and other instructors), especially in course outlines and course syllabi, but also through grading rubrics for assignments and exams.
- An instructor should inform students of their progress in the course through the timely return of graded assignments and exams.
- An instructor should make every effort to ensure that gender, race, or other biases do not influence grading.
- An instructor should address grading practices and grading policies in annual self-evaluations and promotion documents. Similarly, deans and department heads should look for evidence of deliberate grading policies and practices as part of evaluation of teaching.
- Lower-level courses, especially those serving as a prerequisite for upper-level courses, should align grades to predict student success in upper-level courses.

- Many departments have successful models of organizing simultaneous, multi-section courses and monitoring consistency during and between terms, ranging from strict coordination of content, assignments, and exams between all sections to looser coordination through pre-, mid-, and post-term debriefing sessions.
- Deans and department heads are encouraged to foster dialogue about grading practices and grading policies at the college and department levels. For instance, a faculty meeting might be devoted to addressing any observed or perceived issues regarding consistency of grades, especially between related courses. More formally, a dean might request annual reports from department heads on trends and anomalies in final course grades and plans to address identified issues.
- Innovative teaching and grading strategies (e.g., contract learning) are to be encouraged and might be discouraged by a single system meant to ensure consistent grading.

References and Resources

Barbara Gross Davis. *Tools for Teaching (Second Edition)*. Jossey-Bass Publishers, 2009. [Chapter 43: Grading Practices; Chapter 44: Calculating and Assigning Grades].

Gerald S. Hanna and William E. Cashin. *Improving College Grading [IDEA Paper No. 19]*. Kansas State University, Center for Faculty Evaluation & Development, January 1988.

Howard R. Pollio and W. Lee Humphreys. Grading Students. *New Directions for Teaching and Learning*, 1988(34):85–97, 1988.

Robert Rotenberg. *The Art & Craft of College Teaching: A Guide for New Professors and Graduate Students (Second Edition)*. Left Coast Press, 2010. [Part 17: Effective Evaluation of Student Achievement; Part 10: Assessing Student Learning].

Marilla Svinicki and Wilbert J. McKeachie. *McKeachie's Teaching Tips: Strategies, Research, and Theory for College and University Teachers (Fourteenth Edition)*. Wadsworth Publishing, 2013. [Chapter 10: Assigning Grades: What Do They Mean?; Chapter 7: Assessing, Testing, and Evaluating: Grading Is Not the Most Important Function].

Barbara E. Walvoord and Virginia Johnson Anderson. *Effective Grading: A Tool for Learning and Assessment*. Jossey-Bass Publishers, 1998. [Chapter 6: Calculating Course Grades; Chapter 5: Establishing Criteria and Standard for Grading; Chapter 7: Communicating with Students About Their Grades].