



# Final Exam Weighting as Part of Course Design

## ABSTRACT

The weighting of a final exam or a final assignment is an essential part of course design that is rarely discussed in pedagogical literature. Depending on the weighting, a final exam or assignment may provide unequal benefits to students depending on their prior performance in the class. Consequently, uncritical grade weighting can discount student learning, by ensuring that improved mastery of material at the semester's end is not reflected in the course grade. Problems related to several common final exam weights are explored, as are potential solutions to unequal student outcomes made possible by uncritical grade weighting. Ultimately, this essay argues that choosing a weight for a final exam or a final assignment determines what types of student success ought to be possible in the class; therefore, instructors should assign exam weights intentionally, being fully aware of the potential benefits and problems of the weights that they choose.

## KEYWORDS

final exam weighting, assignment weighting, course design, prefinal grades, grading plateaus

## INTRODUCTION

Over the last thirty years, a significant body of scholarship on teaching and learning in the college classroom has redefined effective college teaching. Central to such scholarship is a concern with student learning and classroom participation through innovative uses of new technology, transparency in grading, and strategic use of learning objectives. These reforms have affected almost all aspects of collegiate pedagogy.

However, a nuanced discussion of grade weighting as a pedagogical tool is missing from this new work. For example, O'Brien, Millis, and Cohen address almost every aspect of course design in a learning-centered context, except for grade weighting, merely stating that "you should discuss how you determine the final grade" in the syllabus (O'Brien et al, 2008, p. 99). Barbara Gross Davis does not consider grade weighting except to note that "the final should count for no more than one-third of the course grade" because a heavily-weighted final might encourage cramming (Davis, 1993, p. 288). While others describe different systems of grade weighting (Walvoord & Anderson, 1998, p. 93-97), or acknowledge that exam weighting can affect student grades (Milton, Pollio, and Eison, 1986, p. 44), none of these authors discuss the weighting of exams in relation to learning. The only studies to address exam weighting focus on the weighting of individual questions within exams, not the weighting of exams within a course (Feurman & Weiss, 1973; Bhaskar & Dillard, 1983).

The lack of scholarship on assignment weighting in general, and on final exam weighting in particular, is mirrored by the lack of scholarly consensus on the relationship between grades and learning. As is commonly acknowledged, grades—whether numerical scores attached to particular assignments, the abstract course grades derived from a number of individual scores, or even more abstractly, GPAs of the type used in the United States—are not an ideal measure of student learning.

Alfie Kohn (1999) argues that the seemingly objective aspects of grades actually reflect instructor choices, rather than student ability: “Even the score on a math test is largely a reflection of how the test was written.” Rita Berry (p. 237, 2008), on the other hand, views grades as “the outcome of learning” which reflects “students’ achievement” (p. 125), although she acknowledges that students can become obsessed with grades to the detriment of their education (p. 125-26). One recent study suggests that GPA does not accurately reflect learning (Campbell & Cabrera, 2014); another (Mostrom & Blumberg, 2012) proposes that active learning strategies can improve students’ understanding of the material and thus their grades, which indirectly feed into their GPAs. Perhaps the issue is moot: Students can even learn simply from being tested, regardless of their grade (Glass, Ingate, & Sinha 2013, p. 224-25).

This essay is not designed as a contribution to the debate over the effectiveness of grades, but rather as an exploration of the practical reality of grading systems as they are used today. Although grading systems may be imperfect, many instructors are expected (even required) to use them, so the issue is how to use grades in the most effective way to encourage student learning (Allen, 2005, p. 218-19). Recent scholarship endorsing grades largely focuses on grading chiefly as a way of communicating feedback about the quality of student work (Walvoord & Anderson, 1998; Marzano, 2000; Berry, 2008). Transparency between instructors and students regarding grading criteria and practice are essential (Aufrecht, 1997, p. 50-51). Bringing grading strategies in line with learning objectives is a good method for encouraging students to focus on learning (Farias, Farias, & Fairfield, 2010, p. 339).

With these points in mind, the present essay examines grade weighting in light of its potential relationship to learning. This article, therefore, provides tools to improve the standard grade weighting practice, in which instructors assign weight to an assignment based on their perception of its importance. For although grades are supposed to reflect growth or proficiency, it’s possible to assign weights to final exams and assignments that fail to reflect students’ abilities: examples include exams that are weighted so lightly that they make no difference to any student’s grade, and exam weights that offer drastically unequal chances at grade improvement, regardless of the students’ scores on the exam itself. Such grading policies effectively discount student learning, and create a phenomenon similar to grade inflation, but in reverse: At some exam weightings, most course grades can only go down or stay the same. Assigning weights to grades, therefore, provides a way for instructors to make key decisions about how assignments should reflect student learning or proficiency. Consequently, this essay analyzes grade weighting systems, demonstrating their potential problems and inequities, and proposes methods for avoiding situations in which grades become decoupled from learning.

Further side effects of grade weighting remain to be explored, including its potential influence on student motivation and performance; there is no room to examine this topic in detail here. However, existing scholarship on low-stakes exams suggests that students’ motivation and performance might suffer (Wise, 2009; Napoli & Raymond, 2004), while previous work on high-stakes exams indicates that stress could also be a side effect (Wolf and Smith, 1995).

While this study was chiefly designed with final exams in mind, the essential concepts related to grade weighting discussed here are readily transferable to the weighting of final projects, assignments, and essays. The discussion of grading plateaus may also prove relevant for instructors who prefer to assign portfolios or final performances. As the core of this analysis is based on the mathematical possibilities for grade improvement implicit in various final exam/assignment weights, the results should be applicable to many teaching situations, regardless of subject matter or student demographics.

As will be shown below, it is difficult to create weighting systems in which all students’ performance on the final exam matters equally (to some extent, the importance of a final exam or

assignment for a student's course grade depends on previous student performance in the class). However, it is possible to minimize the damage of uncritical assignment weighting.

## RATIONALE

The following observations are guided by a simple principle: A final exam or assignment should allow the grades of a reasonable percentage of students to improve. The percentage of students affected and the amount of improvement possible can vary on the course taught and by discipline.

Room for improvement in student grades on the final graded activity (whether an exam, an essay, a project, presentation, etc.) in a course is generally desirable, because grades should reflect student performance in a meaningful way. If enough students are unable to improve their grades through a final graded activity, the grading system may have ceased to reward growth or to reflect proficiency, and might be working against the instructor's objectives for the course. (There are several exceptions: students who have already earned the highest possible grade in the course will not be able to improve their grades; and students who are failing the course might similarly be unable to improve their course grade in some contexts, no matter how well they score on the final exam.)

The problem is that some exam weights make grade improvement impossible for all but a few students, while others make the bar for grade improvement so low as to greatly reward only marginal improvement. Although such grading systems may serve perfectly well in certain pedagogical contexts, the general scholarly silence on grading weighting as a tool of course design suggests that grade weighting systems are being used uncritically.

Ultimately, the argument here is not that instructors should use one particular grade weighting system over another, but rather that instructors should use these weightings intentionally and with full awareness of the consequences. It's easy to view the mathematical workings of grade calculation as objective and impartial; in reality, the "math of the course" is something that instructors create when they assign weights to assignments in a class. Steven Aufrecht perceptively notes, "Those who argue that the evaluation process is objective and that the evaluator has no discretion conceal, sometimes even from themselves, the asymmetrical balance of power in this relationship" with the students being evaluated (Aufrecht, 1997, p. 51). As will be shown below, a weighting system structures and limits the range of graded student outcomes and can weaken the connection between grades and student achievement. This is true regardless of whether or not the designer is aware of the system's impact on students' grades.

Assignment weights, when used intentionally, reveal the instructor's specific priorities regarding the importance of various assignments and the types of graded student outcomes that should be possible. One instructor might think it appropriate for a student whose performance in class has been undistinguished at best to be able to reach the next grade level by acing the final, because a high score on the exam indicates increased mastery of course content; another might offer a low-stakes final of less than 10% because the deciding work of the course occurred during the semester and cannot be measured by any test. As will be shown below, setting the weight of a final exam has far-reaching consequences on the ways that student learning is reflected in the final course grade.

## METHODOLOGY

The following analysis is not comprehensive, as a variety of grading scales are used around the world. However, the issues addressed here are implicit in any type of grading system that converts numerical scores into ranked categories. This essay will focus on grading systems used in the United States, as these are the systems with which the author is most familiar: the pure 10-point system (A =

90–100, B = 80–89, etc.), the 10-point system with pluses and minuses (A = 93–100, A- = 90–92, B+ = 87–89, B = 83–86, etc.), and the 7-point system (A = 94–100, B = 87–93, C = 80–86, D = 70–79, etc.).

The basic mathematical procedure consists of weighting the student’s prefinal grade (their grade before the final exam) and the final exam grade to generate the course grade. This discussion does not consider curved grading in any way. Calculations can be made with either raw points or abstract, percentage-based grades.

Here is a brief example, using percentage-based grades. Our hypothetical student, Andrea, has a prefinal score of 90% and takes a final worth 10% of the course grade. The prefinal score is worth 90% of the course grade. Consequently, the prefinal grade is multiplied by 0.9, giving a score of 81 points out of 100 ( $90 \times 0.9 = 81$ ); Andrea earns a 76% on the final exam, a score that converts into 7.6 points out of ten ( $76 \times 0.1 = 7.6$ ). Thus her numerical course grade is 88.6%.

In the example just mentioned, Andrea would need at least 90% on the final exam in order to secure an A on a ten-point grading scale (to move from 81/90 before the final to 90/100). This example exposes a basic rule of exam weighting: for a student’s course grade to improve, the student’s final exam performance must exceed their prefinal grade. Thus (still assuming a final exam worth 10% of the final grade), a student with a prefinal score of 75% (67.5 points), needs to earn at least a 76% on the exam (7.6 points) to raise their numerical course grade at all (to 75.1 points), although this would probably make no difference under most grading scales.

The relationship between Andrea’s prefinal grade and her final course grade changes depending on the grading scale used. If Andrea’s instructor uses the 10-point system with pluses and minuses, her grade falls from an A- (90% prefinal) to a B+ (88.6% final course grade). If the pure 10-point system is used, then her grade changes from an A to a B. If the 7-point system is used, then her prefinal and final grades remain the same (B).

A points-based version of the same process might look like this: Arjun takes a course that uses an assignment weighting system worth 600 points; the instructor assigns letter grades based on the 7-point scale (A = 564 and higher, B = 522–563, C = 480–521, etc.). The final assignment is worth 120 points. Arjun has earned 394 of the 480 points available before the final assignment: his prefinal grade is 82%. He receives a 96/120 (80%) on the final assignment for a final course grade of 490/600 (81.6% = C).

One might suppose that points-based grading systems undercut the weighting of assignments as all points are equal in weight, and instructors do not explicitly assign percentage-based weights to any assignment (such as making a final exam worth 20% of the course grade). This is an error, however; while all points are indeed equal to each other, instructors assign weights by making specific exams or assignments worth *more points*. Further, all point-based systems can easily be converted into percentages and the same issues discussed below—plateaus, ceilings, etc.—still apply to them.

#### AN EXTREME CASE: 5% EXAMS

Perhaps no assignment weighting is a better example of the disconnect between grades and student achievement than a final exam worth 5% of the course grade. At this weight, grade improvement is impossible for almost all students, and any improvement in the student’s score is likely to be discounted by the grading system. A student with a prefinal grade of 65%, for example, can earn a maximum course grade of 66.75%—which hardly makes a difference to a student’s final grade under most grading scales.

Borderline grades in a 10-point system (69, 79) have a marginal chance of improvement: a student with a 69% prefinal grade would need to score at least 89% on the exam to raise their overall course grade to a 70%, while a student with a 79% prefinal grade would need at least a 99% on the exam

to raise their overall course grade to 80%. Meanwhile, a student with a prefinal grade of 89% can only reach a 90% if the instructor rounds grades up. Among students with failing scores, only those with 58% and 59% prefinal scores have a chance of reaching 60% (D).

In a 7-point system, the 5% exam is unlikely to offer any possible grade change at all, except to failing students. The highest-scoring students in the B range (93% prefinal) and the highest-scoring students in the C range (86% prefinal) are unable to reach the next letter grade without extra credit. Meanwhile, the highest-scoring students in the D range (79% prefinal) can only reach 80% (C) by scoring 99% on the final. The highest-scoring students in the F range (69% prefinal), on the other hand, can reach a D if they score an 89% or better.

A final exam worth 5% of the grade makes so little difference to the students' grades that there can be limited penalties for skipping the exam, especially in a 10-point system. For example, a student with a prefinal grade of 75% would receive a course grade of 71.25% if they skipped the final. This effectively means that under a plain ten-point scale (A = 90–100, B = 80–89, etc.), there is often no penalty for skipping the exam. At the same time, students with borderline grades face more severe penalties: if students with a prefinal grade of 82% skip the final, their course grades will drop to 77.9%. However, if these students took the final—and failed it with a score of 50%—their course grade would only drop to 80.4, which makes no difference to their course grade under the 10-point scale, with or without plus and minus grades. So students on the borders between grades would feel compelled to attempt the final if they understood the odds (they might not, however, without a clear explanation from the instructor).

A 5% final exam therefore makes it difficult for grades to reflect any improvement in content mastery or any growth in student knowledge by holding students to an impossibly high standard. Students will have little motivation to put forth their best effort if they understand the odds outlined above. A student with a 77% prefinal grade who receives 100% on the final exam, yet still receives a C in the course, is likely to feel that their effort to learn has been devalued.

One way that a 5% exam might be pedagogically effective is in a proficiency-focused course which values mastery or technical ability. Students might be required to earn a certain score on the final exam to pass the course (regardless of the exam's weight); this would essentially shift the importance of the exam away from its weight and toward the students' scores.

A 5% final exam may seem unrealistic or out of touch with common pedagogical practice. However, Fink's handbook for college teachers approvingly lists a course weighting system in which a final exam is worth 10/150 points (6.66% of the course grade), which is only marginally more effective than a final exam worth 5% of the course grade (Fink, 2013, p. 158). Little explanation is given for this weighting. Our hypothetical student with a 77% prefinal would still get a C if the course was graded on a 10-point scale.

## PLATEAUS AND SYSTEMATIC DISADVANTAGES IN GRADE WEIGHTING

From the examples above, it should be clear that, regardless of what students have learned, the weighting of a final graded activity can make it practically impossible for some of them to improve their grades at the end of the course. Grades do not simply reflect learning or mastery; instructors choose to give grades that power.

A comparative approach to exam weighting demonstrates this point. Figures 1 and 2 show how a student's course grade is altered by the instructor's decision to assign a weight to the final exam as part of a larger grading system. Thus, a student with a 79% prefinal score who wants to reach an 80% in the course needs to score a 99% on a final weighted at 5%, but only 81% on a final weighted 50%, regardless

of whether the course is graded on a 7-point or 10-point scale. An exam grade that improves the overall course grade can thus vary from two to eighteen points above the prefinal score, if the instructor so wishes. These are extremes and not norms, but they illustrate the hidden consequences of final exam weighting.

Figure 1

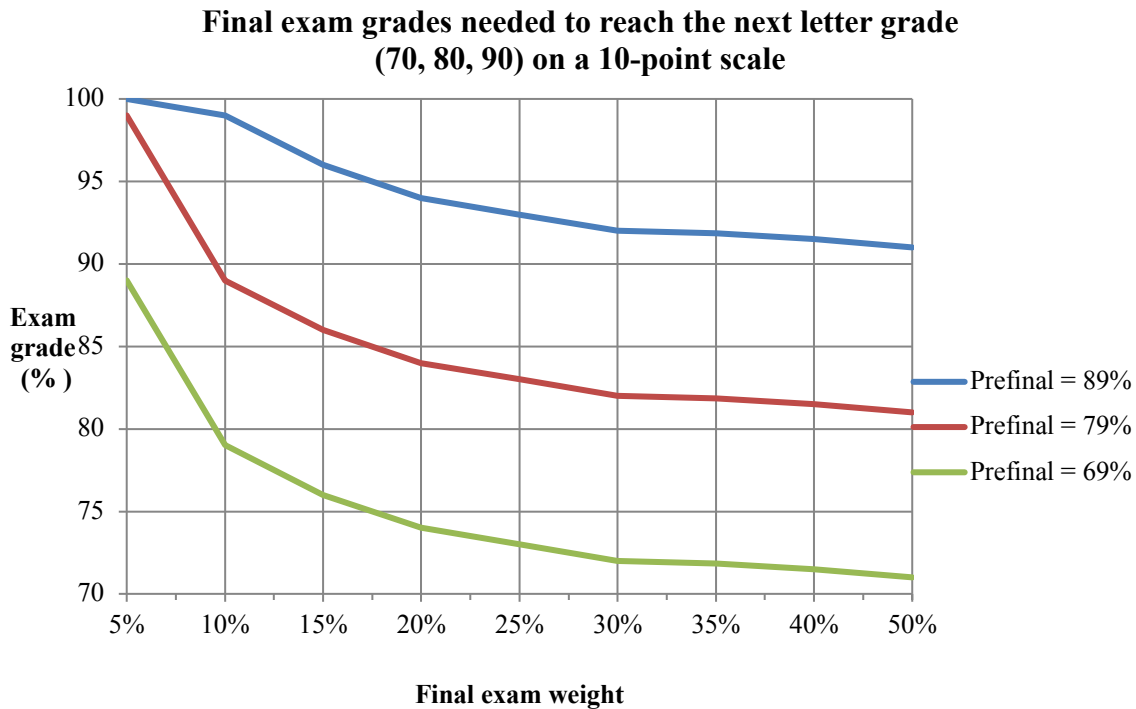
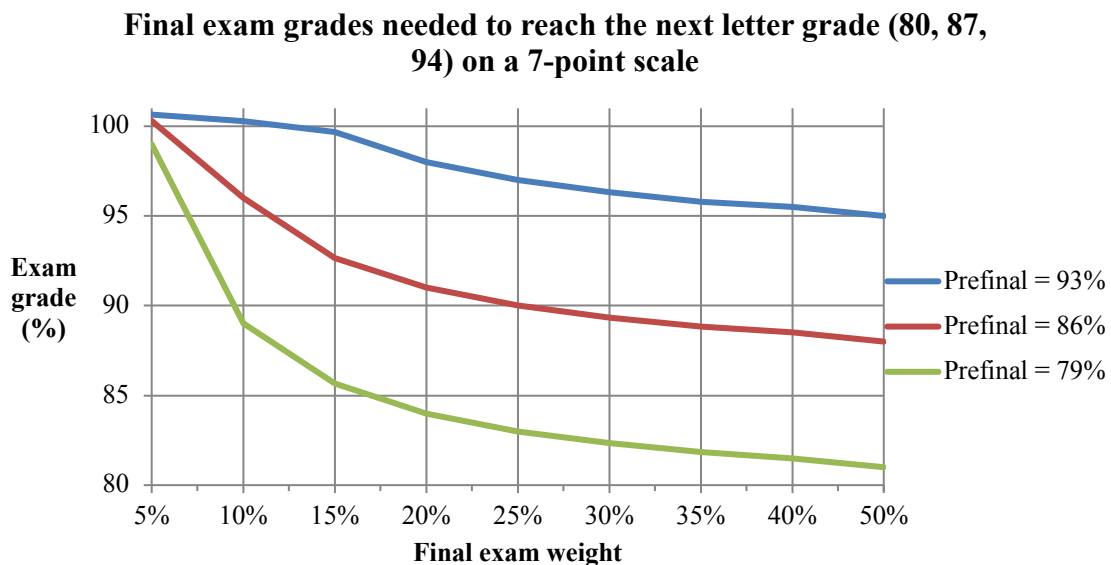


Figure 2



As already established, a final exam weighted at 5% makes it essentially impossible for most students to improve their grades, and is, without measures to restore its relevance, effectively a “low-stakes” exam, to borrow Cole and Osterlind’s phrase for “any exam that has no meaningful consequence to the test taker” (Cole & Osterlind, 2008, p.120). If students accurately understand the effect that the exam will have on their course grade, then they are likely to have low motivation to succeed (Wise, 2009; Napoli & Raymond, 2004). However, unless professors are transparent with students about the limited potential for grade improvement, it is likely that at least some students will still view the exam as a high-stakes exam, regardless of its weight. Due to films such as *Bill & Ted’s Excellent Adventure* (1989), students sometimes expect that excellent performance on a final exam or assignment will make amends for all previous unsatisfactory work, regardless of exam weight.

Similarly, Figures 1 and 2 suggest that it makes very little difference in grading terms whether the final exam is worth 30–50% of the course grade—the amount needed to improve student grades is essentially the same. The real difference is that a heavily-weighted final might have a drastic effect on student stress (Wolf & Smith, 1995, p. 229-30), although student stress cannot be wholly attributed to the stakes of an exam in such a case (Sloane & Kelly, 2003, p. 13-14). Consequently, there is at least the possibility that grading systems might potentially have detrimental effects on students’ emotional health.

Regardless of whether or not stress is a definite result of exam weighting, Figure 1 suggests that there are diminishing returns on offering exams weighted more than 30%. The descending lines start flattening between 30% and 50%; similar plateaus occur in Figure 2 (the 7-point system), although these start at 35%. Because of the plateau above 35%, a final exam or assignment weighted between 10% and 30% seems most likely to insure that student work can reflect achievement by allowing some room for grade improvement.

Yet even within this range, there are problems, as can be seen from **Table 1**, which lists the maximum course grades for each prefinal grade at various final exam weightings. Students with the highest prefinal grades have the least chance of improving their course grade, regardless of the weight of the exam or their mastery of course content.

In a 10-point system, an exam worth 10% of the course grade offers unequal opportunities for grade improvement. The top three-tenths of students with a prefinal grade of D (67–69%) in the course can reach 70% (the next letter grade) if they earn the maximum score on the final exam. However, only the top two-tenths of students with a prefinal grade of C (78–79%) can reach 80% (the next letter grade) if they earn the maximum score on the final exam. And finally, only the top tenth of students with a prefinal grade of B (89%) can reach 90% (the next letter grade) if they earn the maximum score on the exam. A 10% final exam, offered in a 10-point system, does therefore reflect student achievement, but only for students with one of six passing prefinal grades (67%, 68%, 69%, 78%, 79%, and 89%) and four failing prefinal grades (56%, 57%, 58%, and 59%). Regardless of how well everyone else performs, their grades will either stay the same or go down. In a 7-point system, the benefits are generally narrower. At 10%, only students with passing prefinal scores of 78%, 79%, and 86% can reach the next letter grade (as well as 85% and 93% if the instructor rounds up).

Increasing the exam weight in either a 7- or 10-point system expands the potential benefit: at 15%, students with a 65–69 prefinal score can reach the next letter grade in a 10-point system, as can students with a 64, 76, and 88 if the instructor rounds up scores. At the same time, students with prefinal scores of 60–63, 70–75, and 80–87 are essentially unable to improve their course grades unless a +/- grading scheme is being used. However, a final exam weighted at 15% does avoid the motivation issues noted above with a 5% exam: most students cannot score a 50% on the exam without significantly affecting their overall course grade. At 25%, the opportunity for reaching the next letter grade has

expanded to students with 65–69, 74–79, and 87–89 prefinal grades (73–79 and 85–89 if the instructor rounds up grades).

These examples suggest that higher exam weights provide more room to reward the type of student learning that can be reflected through an exam score, but there are problems with this weighting as well. At 25%, a student with a prefinal of 65% can score a maximum course grade of 73.75%, or 8.75 points above the prefinal. At the same exam weight, a student with a prefinal of 89% can only score a maximum of 2.75 points above the prefinal grade. This phenomenon is observable regardless of which exam weight or grading scale is used.

Ultimately, whatever the exam weighting, students with lower prefinal grades are more likely to profit than students with higher prefinal grades, whose letter grades are effectively limited by a ceiling effect. Thus, the grading systems outlined above only reflect *some* students’ learning and are not necessarily accurate reflections of student learning across a class as a whole. Imagine two students who both score 100% on a final exam or activity weighted at 15% of the course grade in a 10-point system: one student has a 77% prefinal, and thus earns an 80% (B) in the course, while the student with an 87% prefinal stays at 89% (B). Under this grading system, students with weaker prior performance in the course stand to benefit more than those with higher demonstrable mastery of course content. The ceiling effect does not reward two equal exam scores with an equal change to the course grade. While one might say that this is simply “the math of the course” at work, such a view hides the informed instructor’s agency in choosing to allow the ceiling effect to work this way: It is possible to choose a different weighting system to avoid this type of effect.

There is also a hidden benefit to ceiling effects. While ceiling effects limit the potential for grade improvement, grade weights also create floors below which grades may not drop, depending on prior student performance. In a final exam weighted at 10%, for example, a student with a prefinal grade of 80% is guaranteed to receive at least 72% in the course overall. However, as noted above, if a floor is too high, then the stakes of the exam might become too low, and students might lose their motivation to take the test. Thus it might be worth adding an additional incentive to a lower-stakes exam, to raise the stakes slightly. At the same time, high floor effects might reduce student stress over their grades; some instructors might, therefore, prefer to balance ceiling effects with floor effects.

**Table 1** Maximum course grade scores, given prefinal grades and final exam weights

| PREFINAL GRADES | FINAL EXAM WEIGHT<br>5% | FINAL EXAM WEIGHT<br>10% | FINAL EXAM WEIGHT<br>15% | FINAL EXAM WEIGHT<br>20% | FINAL EXAM WEIGHT<br>25% |
|-----------------|-------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 65              | 66.75                   | 68.5                     | 70.25                    | 72                       | 73.75                    |
| 66              | 67.7                    | 69.4                     | 71.1                     | 72.8                     | 74.5                     |
| 67              | 68.65                   | 70.3                     | 71.95                    | 73.6                     | 75.25                    |
| 68              | 69.6                    | 71.2                     | 72.8                     | 74.4                     | 76                       |
| 69              | 70.55                   | 72.1                     | 73.65                    | 75.2                     | 76.75                    |
| 70              | 71.5                    | 73                       | 74.5                     | 76                       | 77.5                     |
| 71              | 72.45                   | 73.9                     | 75.35                    | 76.8                     | 78.25                    |
| 72              | 73.4                    | 74.8                     | 76.2                     | 77.6                     | 79                       |



|    |       |      |       |      |       |
|----|-------|------|-------|------|-------|
| 73 | 74.35 | 75.7 | 77.05 | 78.4 | 79.75 |
| 74 | 75.3  | 76.6 | 77.9  | 79.2 | 80.5  |
| 75 | 76.25 | 77.5 | 78.75 | 80   | 81.25 |
| 76 | 77.2  | 78.4 | 79.6  | 80.8 | 82    |
| 77 | 73.15 | 79.3 | 80.45 | 81.6 | 82.75 |
| 78 | 79.1  | 80.2 | 81.3  | 82.4 | 83.5  |
| 79 | 80.05 | 81.1 | 82.15 | 83.2 | 84.25 |
| 80 | 81    | 82   | 83    | 84   | 85    |
| 81 | 81.95 | 82.9 | 83.85 | 84.8 | 85.75 |
| 82 | 82.9  | 83.8 | 84.7  | 85.6 | 86.5  |
| 83 | 83.85 | 84.7 | 85.55 | 86.4 | 87.25 |
| 84 | 84.8  | 85.6 | 86.4  | 87.2 | 88    |
| 85 | 85.75 | 86.5 | 87.25 | 88   | 88.75 |
| 86 | 86.7  | 87.4 | 88.1  | 88.8 | 89.5  |
| 87 | 87.65 | 88.3 | 88.95 | 89.6 | 90.25 |
| 88 | 88.6  | 89.2 | 89.8  | 90.4 | 91    |
| 89 | 89.55 | 90.1 | 90.65 | 91.2 | 91.75 |
| 90 | 90.5  | 91   | 91.5  | 92   | 92.5  |
| 91 | 91.45 | 91.9 | 92.35 | 92.8 | 93.25 |
| 92 | 92.4  | 92.8 | 93.2  | 93.6 | 94    |
| 93 | 93.35 | 93.7 | 94.05 | 94.4 | 94.75 |

### THREE PROBLEMS, FOUR COUNTERARGUMENTS, AND TWO SOLUTIONS

To summarize, three problems have emerged from this analysis of final exam weighting:

(1) If a final exam is worth too little, students' scores on the exam make almost no difference to their course grades and fail to reflect their learning.

(2) Students do not have equal opportunity to improve their grades through their performance on final exams, meaning that the grading system may have become detached from the instructor's pedagogical goals. Students with low prefinal grades have a higher chance of improving their course grades than do students with high prefinal course grades.

(3) A heavily-weighted final exam, while it does increase the opportunity for grade improvement, might potentially increase student stress and thus be counterproductive. Heavily-weighted exams still benefit students with lower prefinal grades more than students with high prefinal grades.

There are several possible counterarguments to points (1) and (2):

(1) A low-stakes exam might be a useful pedagogical tool even if the exam has little effect on the course grade. It could provide an environment for low-stress application of class concepts, and be quite

useful if the instructor provides some kind of additional motivation for students to do well. It would also work well if the class focused on proficiency as opposed to growth.

(2a) One could argue that the inequalities of exam weighting are acceptable as long as the students understand them beforehand.

(2b) One could argue that if the students want to avoid the unequal chances of this system, they should strive to do their best work before the final to benefit from floor effects; this, however, requires that the instructor work vigilantly to assure student understanding of the biases of the grading system.

(3) A proficiency-based course might use a heavily-weighted final to communicate the importance of the assignment and mastery of the material at the end of the semester.

Because effective teaching can take many forms, these counterarguments are worth considering seriously, and they may be convincing in certain pedagogical contexts. Yet they do not address my central concern with the current system of exam weighting: that various exam weights create ceiling effects that make it too easy for students with low prefinal grades to improve their scores, while denying the same opportunity to higher-scoring peers.

One solution, which works for comprehensive final exams, is to decouple final examinations from course grades, counting them essentially as elaborate extra-credit assignments. (This doesn't work well if the final is not comprehensive, because it makes the assessment of new material optional). In this scenario, all course work before the final would equal 100% of the course grade, and students would have the option of not taking the final exam if they were satisfied with their course grade. Successful work on the final exam would grant students an extra credit score to be determined by the instructor. An example of such a system is shown in Table 2.

**Table 2 A possible rubric for treating a final exam as an extra credit assignment**

| FINAL EXAM SCORE | BONUS TO THE COURSE GRADE |
|------------------|---------------------------|
| 96–100           | +4 percentage points      |
| 91–95            | +3 percentage points      |
| 85–90            | +2 percentage points      |
| 79–84            | +1 percentage point       |
| 0–78             | No bonus                  |

This system has several benefits:

(1) All students' grades on the final exam have the same effect on their course grade, regardless of their prefinal grade. Consequently, all students' learning can have the potential to affect their overall course grade, depending on how much extra credit the instructor makes available, and the ceiling effect can disappear.

(2) Students who have a borderline grade can tell at a glance what grade they need on the final exam to reach their desired course grade.

(3) Students who have the highest possible passing grades or the lowest possible failing grades will generally have no reason to take the final exam, unless it serves a particular pedagogical purpose, such as assessing student retention and mastery of course content. Not taking the exam reduces students' workload and might also reduce their stress level. Not having to grade as many exams reduces faculty workload as well. Thus this system avoids the pointless labor of the current system, in which

students who have demonstrated their mastery of the material throughout the semester often still have to take a comprehensive exam, solely to keep the grade that they have already earned throughout the semester.

(4) This system can be easily customized to grant as many or as few bonus points as are needed in accordance with the instructor's pedagogical goals.

However, if this system is adopted, then all the problems laid out with final exam weighting in this article will apply to the last graded assignment of the course. For the reasons outlined above, the last assignment before the final exam should probably still be weighted somewhere in the 10–25% range. Otherwise, it too, could have such a marginal effect on the students' course grades that it might cease to reflect their learning.

An alternate method of assigning extra credit points to minimize the ceiling effect on a comprehensive final exam involves using Hake's normalized gain, a measure that is commonly used to measure the effectiveness of force concept inventories in physics courses (Hake, 1998). Simply put, the measure defines gain (G) as follows:

$$G = \frac{\text{postscore \%} - \text{prescore \%}}{100 - \text{prescore \%}}$$

With a comprehensive extra-credit final, one can substitute the final exam score for the postscore and the preliminary course grade (the course grade the student would have if they didn't take the final exam) for the prescore:

$$G = \frac{\text{final exam score \%} - \text{preliminary course grade \%}}{100 - \text{preliminary course grade \%}}$$

This procedure measures gain in student achievement between the course as a whole and the final exam. Gain can then be correlated with a set number of extra credit points, at the instructor's discretion, as above, with  $G = 1.00$  representing the highest possible gain. This method, however, has the problem of being somewhat obscure and difficult to explain to students, limiting the transparency of the measure.

The ceiling effect in non-comprehensive final exams remains a problem. An extra-credit exam seems ineffective (it creates the possibility that students might never be tested on some of the course material), but might be effective if other means of assessment were also in place. Probably the simplest way of redressing the inequities of the ceiling effect with a non-comprehensive final is for the instructor to provide generous opportunities for extra credit so that equal opportunity is restored for students with "borderline" grades. For example, a 10% final in a 10-point system allows students prefinal grades of 77%, 78%, 79%, 88%, and 89% to reach the next letter grade: so an effective way to weaken the ceiling effect would be to make an extra-credit assignment worth 1–2% of the course grade, so that students at 87% would have similar chances to students at 77%. Note that this doesn't actually erase the ceiling effect; the student with an 87% still has a lower chance of reaching 90% than a student with 77% does of reaching 80%. It is, however, at least an attempt at equalizing the importance of student scores on final exams and reflecting all student learning somewhat equally in the course grade.

## CONCLUSIONS

Weighting the final exam thus becomes a balancing act for instructors: Should students with a 74% prefinal be able to reach a course grade of 80%? Is it a fair representation of student learning if a student with a prefinal of 88% might not be able to reach the next grade level, even if she scores 100% on

the final? Is it fair that, given the same exam at the same weight, a student with a prefinal of 78% can reach 80% (and the next grade level)? Should the final exam have a lower weight to avoid potential stress? But if the exam has too little weight, will it make any difference in the students' course grades? And if the grades do not reflect learning, retention, or mastery, then how does the final exam or assignment serve the instructor's pedagogical aims?

Ultimately, exam weights can have significant and often unfair effects on student grades if they cease to reflect student learning or proficiency. Consequently, instructors must maintain transparency with students about the effect that exam weights can have on their grades. There are a number of ways to use exam weighting effectively, but instructors must be aware of the power of grade weighting and make their choices intentionally.

*Matthew Franke is a Master Instructor in Music History at Howard University in Washington, D.C.*

## REFERENCES

- Aufrecht, S. E. (1997). Grading as performance evaluation. *Journal of Public Administration Education* 3(1), 43-57.
- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House* 78(5), 218-23.
- Berry, R. (2008). *Assessment for Learning*. Hong Kong: Kong University Press.
- Bhaskar, R., & Dillard, J. F. (1983). Using cognitive science to assign test weights. *Instructional Science* 12(4), 375-82.
- Campbell, C. M., & Cabrera, A. F. (2014). Making the mark: Are grades and deep learning related? *Research in Higher Education* 55, 494-507.
- Cole, J. S., & Osterlind, S. J. (2008). Investigating differences between low- and high-stakes test performance on a general education exam. *Journal of General Education* 57(2), 119-30.
- Davis, B. G. (1993). *Tools for Teaching*. San Francisco: Jossey-Bass.
- Farias, G., Farias, C. M., & Fairfield, K. D. (2010). Teacher as judge or partner: The dilemma of grades versus learning. *Journal of Education for Business* 85, 336-42.
- Feuerman, M., & Weiss, H. (1973). A mathematical programming model for test construction and scoring. *Management Science* 19(8), 961-66.
- Fink, L. D. (2013). *Creating Significant Learning Experiences: An Integrated Approach to Designing College Courses* (2nd ed.). Somerset, US: Jossey-Bass.
- Glass, A. L., Ingate, M., & Sinha, N. (2013). The effect of a final exam on long-term retention. *Journal of General Psychology*, 140(3), 224-41.
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics* 66 (1), 64-74.
- Kohn, A. (1999). From grading to de-grading. *High School Magazine* (March issue). Retrieved from <http://www.alfiekohn.org/article/degrading-de-grading/>
- Marzano, R. J. (2000). *Transforming Classroom Grading*. Alexandria, VA: Association for Supervision and Curriculum Development. Retrieved from <http://www.ascd.org/publications/books/100053/chapters/What-Are-Grades-For%C2%A2.aspx>
- Milton, O., Pollio, H. R., & Eison, J. A. (1986). *Making Sense of College Grades*. San Francisco: Jossey-Bass.
- Mostrom, A. M., & Blumberg, P. (2012). Does learning-centered teaching promote grade improvement? *Innovative Higher Education* 37, 397-405.
- Napoli, A. R., & Raymond, L. A. (2004). How reliable are our assessment data?: A comparison of the reliability of data produced in graded and un-graded conditions. *Research in Higher Education* 45(8), 921-29.
- O'Brien, J. G., Millis, B. J., & Cohen, M. W. (2008). *The Course Syllabus: A Learning-Centered Approach* (2nd ed.). San Francisco: Jossey-Bass.
- Sloane, F. C., & Kelly, A. E. (2003). Issues in high-stakes testing programs. *Theory Into Practice* 42(1), 12-17.
- Walvoord, B.E., & Anderson, V.J. (1998). *Effective Grading: A Tool For Learning and Assessment*. San Francisco: Jossey-Bass.
- Wise, S. L. (2009). Strategies for managing the problem of unmotivated examinees in low-stakes testing

programs. *The Journal of General Education* 58(3), 152-66.  
Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education* 8(3), 227-42.



Copyright for the content of articles published in *Teaching & Learning Inquiry* resides with the authors, and copyright for the publication layout resides with the journal. These copyright holders have agreed that this article should be available on open access under a Creative Commons Attribution License 4.0 International (<https://creativecommons.org/licenses/by/4.0>). The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited, and to cite *Teaching & Learning Inquiry* as the original place of publication. Readers are free to share these materials—as long as appropriate credit is given, a link to the license is provided, and any changes are indicated.