A Triage Theory of Grading:

The Good, the Bad, and the Middling[1]

William J. Rapaport

Department of Computer Science & Engineering,

Department of Philosophy, Department of Linguistics,

and Center for Cognitive Science

University at Buffalo, State University of New York

Buffalo, NY 14260-2000

rapaport@buffalo.edu

http://www.cse.buffalo.edu/~rapaport

Abstract

This essay presents and defends a triage theory of grading: An item to be graded should get full credit if and only if it is clearly or substantially correct, minimal credit if and only if it is clearly or substantially incorrect, and partial credit if and only if it is neither of the above; no other (intermediate) grades should be given. Details on how to implement this are provided, and further issues in the philosophy of grading (reasons for and against grading, grading on a curve, and the subjectivity of grading) are discussed.

1 Introduction

In this essay, I present and defend a "triage" theory of grading: An item to be graded should get one of only three grades: *full* credit if and only if it is clearly or substantially *correct*; *minimal* credit if and only if it is clearly or substantially *incorrect*; and *partial* credit if and only if it is *neither of the above.* No other (intermediate) grades should be given. I begin with a discussion of reasons for and against grading; I then turn to the details of the triage theory and its practical implementation, and I close with some remarks on other issues in the philosophy of grading: the subjectivity of grading and grading on a curve.

2 Problems with Grading

> Student to teacher: "If you give us a midterm, you're going to have all of these papers to grade…and I was just thinking…Why not go easy on yourself?" (Batiuk 2004.)

> We are not alone, those of us who indulge in procrastination and get irritable when grading. We are legion. (Clio 2004.)

I hate to grade.

This essay is a contribution to the philosophy of grading. I present a "universal" grading technique, applicable to any discipline. It was inspired by a casual remark made by one of my former professors and has made the task of grading simultaneously easier, more objective, fairer, clearer for the students to understand, and less likely to elicit pleas for "just a few more points" to raise a borderline grade. "One feature of a good grading system is that those measured by it generally regard it as fair and reasonable" (Cohen 2005), even (or especially) those not getting full credit! The underlying insight of the technique to be presented, which many students find "fair and reasonable", is that the work they do is considered to be either good (grade 'A'), bad (grade 'F'), or somewhere in between (grade 'C') and that these are "quantum" units (i.e., there are no grades in between these).

"We have a powerful need to grade" (Hargis 1990: 3). Perhaps it is just an aspect of the human propensity to classify or categorize, or perhaps it is a cognitive imperative with survival value (see, e.g., Mervis & Rosch 1981, Lakoff 1987):

> Adult Student #1: "I think the whole idea of grades in university is ridiculous! We're adults, for crying out loud! We don't care about numbers. We know that true motivation comes from within."
>
> Adult Student #2: "You know, as an arguer I'd give you 8 out of 10."
>
> Adult Student #1: Why only 8?!"
>
> ("Betty" comic strip, November 2002.)

Grading intellectual work is a bit odd when you think about it: Neil Postman notes the "peculiarity" of grading as a "tool" or "technology" for numerically measuring the "quality of a thought", which suggests (presumably falsely) that the measurement is objective and real (Postman 1992: 12–13, 139–140). Nevertheless, no matter how much we might hate to grade and no matter how peculiar it may be, grading students' work is usually required by schools at all levels.

What would happen if we didn't grade? We might resort to what Robert Paul Wolff calls 'criticism':

> The three species of grading are *criticism*, *evaluation*, and *ranking.*
>
> *Criticism* is the analysis of a product or performance for the purpose of identifying and correcting its faults or reinforcing its excellences. …At the elementary level of spelling and syntax…, there is not a great deal of disagreement over what is correct and what is not. When more complex matters of style, argument, and evidence are at stake…, criticism becomes inextricably bound up with intellectual norms which themselves may be matters of dispute. (Wolff 1969: 59.)

We will return to evaluation and ranking in §§6.1 and 6.2. Properly understood, criticism is feedback; it is more akin to the interaction between student and teacher, either the master correcting the apprentice's errors or the two discussing ideas; indeed, as Wolff goes on to say, "*Criticism* lies at the very heart of education" (p. 63).

But criticism can revert to grades, which are often inevitable even in those institutions that claim not to use them. I once taught at the progressive Walden School, in New York City: We did not assign letter or numerical grades to the students. Instead, at the end of each term, we had to write brief paragraphs describing the students' accomplishments. Some students did excellent work, some did good work, some average, some below average, and some did quite poorly. The faculty quickly realized that, in addition to personalized remarks about individual students, there were "boilerplate" remarks for students in each of these categories. If you label them for convenience (say, 'A', 'B', 'C', 'D', 'E'), you find that you have reintroduced grades.

In any case, most, if not all, students *want* some sort of grades; they want to know how they are doing, on either an absolute basis ("Am I a good student?") or a relative basis ("Am I as good a student as others?"). This is no doubt in part due to the prevalence and importance of grades in our academic culture; students have always been graded, so they expect to always be graded.

It may also, in part, be due to a "Dualistic" or "Multiplistic" approach to the nature of knowledge and learning (Perry 1970, 1981). Dualistic students believe that their job in school is to learn Correct Answers[2] to questions posed by Authorities (i.e., by us teachers). Dualistic students see Authorities as teaching by giving them the answers; such students see their job as repeating the answers when asked for them. If they repeat The Correct Answer (there can only be one, of course!), they are good students; if they give "the wrong" answer, they hear us say, "*You* are wrong" and take it as a personal rebuke (even if what we actually said was "*That answer* is wrong"). For such students, grades of 'A' or 'F' make sense; in-between grades don't. After all, the answer is either right, or else it is wrong; there is no room for "partially correct" answers and no understanding of "partial credit" (these terms are seen as oxymorons). Many Dualistic students are attracted to mathematical and scientific subjects because they believe (falsely!) that these subjects always demand clear, right-or-wrong answers: A computer program runs, or else it doesn't; it outputs the correct answer, or else it doesn't; 2+2 always equals 4. And they are similarly put off, or fearful of, less "clear-cut" subjects such as philosophy and literature.

But Dualistic students eventually come to see that there are gray areas, that there are questions whose answers we don't know *yet*: They eventually take the position of "Multiplism". Whereas Dualistic students cannot understand why, for instance, there is more than one theory of morality, or more than one sorting algorithm, or more than one interpretation of a poem ("Which is the correct one?", they wonder), Multiplistic students revel in the multiplicity of different theories, different algorithms with the same input-output behavior, and different literary interpretations.

On the other hand, Multiplistic students are not interested in comparing and evaluating different theories or interpretations, or evaluating input-output–equivalent algorithms in terms of efficiency. That is only appreciated by "Contextually Relativistic" students, who have come to see that all claims must be understood relative to, and in the context of, the evidence that supports them. (I leave for another time the question of how students at this and subsequent Perry positions might view grading.)

Multiplistic students in the early stages of that position see their job as learning how to learn and working hard at it. Grading is a central concern; quantity of work and fairness are seen as the important ingredients of a grade. Thus, such students often complain if they worked for many hours on an assignment poorly done and get a lower grade than their friend who only worked for 15 minutes but who did an excellent job. (Multiplists are further discussed in §§4.2 and 4.4.)

Grades themselves have a dual nature, measuring two things: Numbers (or letters) are assigned to "quality of thought" (to use Postman's phrase), and then ethical or aesthetic values are assigned to the numbers: High grades are good; low grades are bad. Wolff calls this "evaluation" (see §6.1). But not all categorization has to have such ethical value: Red is not better than blue *per se.* A grading system that informs students about their accomplishments in a more-or-less objective way (but see §6.1) might be able to sidestep, if not completely avoid, such an ethically evaluative tar pit.

Both Dualists and Multiplists desire and expect grades. But what should you, the teacher, do when you are faced with grading a highly involved assignment, with many parts and details? Should you take away 1 point for a run-on sentence? Should you take away 1 point for a missing semicolon? (The latter is a classic conundrum for teachers of computer programming, especially because such errors can be found—and automatically corrected—by modern compilers.) And what about essays? Should you give one essay an 'A–' but another, which is only slightly and vaguely poorer, a 'B+'? What is the real difference between those essays (and hence those grades)? And what do you do about the student who wants just a few more points of partial credit (whether or not those few points—perhaps the points for those missing semicolons—will change their grade from 'B+' to 'A–')?

3 The Triage Theory of Grading: Origin and Outline

The Triage Theory of Grading resolves most of these issues. What is "triage"? It is not "the allocation of scarce resources to the 'middle' and none to the top or bottom",[3] as one might expect from an analogy with medical triage in emergencies. Rather, it is merely a method of sorting based on the quality of the items to be sorted; i.e., it is grading. (I say more about the origins of triage in note 7.)

The Triage Theory of Grading is not original with me. I first heard of it in an informal conversation with one of my former professors, Paul Vincent Spade, of the Indiana University Department of Philosophy.[4] Whether or not he intended it seriously, I, and many of my colleagues, have found it quite useful, and my students have found it helpful and fair.

It is based on the following simple observation made after grading freshman philosophy essays: Some are clearly excellent, despite minor problems with grammar, style, argumentation, etc. In general, these students clearly know what they are doing; they pass—give them all grades of 'A'. Other essays are clearly awful in all respects. These students clearly do *not* know what they are doing, or don't care; they fail—give them all grades of 'F'. All the rest of the papers fall somewhere in between these two extremes; they are "average"—give them all grades of 'C'.

The fundamental insight is that, whereas the extremes are clear (it is easy to identify clearly good work and clearly bad work), it is not worth making fine distinctions among the work that is neither clearly good nor clearly bad. Thus, the core idea is to *give only three grades, and none in between.*

Why three and not, say, two? After all, "[i]t is quite possible for a grading system to discriminate between unacceptable and acceptable performances, and yet fail to provide a linear scale of grades along which the various performances can be located" (Wolff 1969: 60). That is, a two-tiered grading scheme might be all that you can have. But I think that there is a middle ground, albeit a large and gray one, between clearly "unacceptable" and clearly "acceptable".

Wolff continues:

> Thus, a connoisseur of violin playing may feel quite confident in judging some performances as excellent and others not, without however having any way of deciding among excellent performances by Heifetz, Millstein, and Oistrakh. The problem is not that they play "equally well," but that beyond a certain level of technical skill and interpretive finesse a choice among them becomes a matter of taste.

Note that here we have triage: unacceptable, at one end; great, at the other; and "technically skilled" in the vast middle.

"*But*…the difference between a great violinist and a bad fiddler *is* a matter of objective evaluation" (p. 60). That is, there are clear differences between top and bottom, but *no* clear differences *within* the top. Thus, there are also probably no clear differences *within* the bottom, *and also no clear differences within the middle.* "[N]o standard, whether pass/fail or letter grades, makes a real delineation between groups of students" (Haladyna 1999: 61).[5] Full credit should be widely separated from no credit, not immediately bordering on it—hence the need for partial credit as a buffer zone. But several refinements and qualifications are possible.

4 The Triage Theory: Details

4.1 Numerical Grading

The first refinement is to grade numerically, not by percentages and not by letters (at least not initially; see §6.1). This has the advantage of not assuming that the grades have any independent or antecedent "meaning": Many students (and teachers) assume, for instance, that 'A' is somehow equivalent to the range 90–100%, 'B' to 80–90%, etc.

(To avoid conflicts, these ranges must be "open" at one end and "closed" at the other end; e.g., 'B' must either include 90% and exclude 80%, or else it must exclude 90% and include 80%. A "lenient" grader will allow the 'A' range to be closed at both ends and all others to be closed only at the bottom: 100% ≤ 'A' ≤ 90%, and 90% < 'B' ≤ 80%, etc. A "stricter" grader would have the 'F' range closed at both ends and all others to be closed only at the top end: 0% ≤ 'F' ≤ 60%, and 60% < 'D' ≤ 70%, etc. We will return to this issue in §6.1.)

I see no rationale for this "classical" mapping of percentages to letters. It is probably a recent invention.[6] What is perhaps the original version—a 100-point system used by mathematicians and philosophers at Harvard in 1837—divided the range (somewhat arbitrarily, it would seem) into: 25 or below, 26–50, 51–74, 75–99, and 100 ("perfect") (Smallwood 1935: 46). Had letters been mapped to these ranges, they clearly would not have matched the "classical" mapping. Indeed, the earliest documented use of letter grades—from Mt. Holyoke College in 1896—had 'A' = "excellent" = 95–100%, 'B' = "good" = 85–94%, 'C' = "fair" = 76–84%, 'D' = "(barely) passed" = 75% (and *only* 75%!), and 'E' = "failed" < 75% (Smallwood 1935: 52).

Numerical grades of the sort that I am about to introduce also have an expository advantage: they allow me to talk about triage grading independently of letters. So, instead of using 'A' for the top grade and 'F' for the bottom grade, I will use the following:

Clearly adequate (full credit)	= 3
Neither clearly adequate nor clearly inadequate (partial credit)	= 2
Clearly inadequate (minimal credit)	= 1

What does 'adequate' mean, however? This will depend on both the subject matter and the type of question or exercise being graded. A simple math problem could have a correct answer, or be solved in an appropriate manner, or its solution by the student might demonstrate clear understanding of the problem. An essay may meet or exceed certain criteria for clarity, exposition, argumentation, creativity, etc. A "skill" (as in a creative writing class, an instrumental music class—recall the discussion in §3, above—or perhaps a programming-language class or a foreign-language class) might be graded on a pre-established level of attainment. (For an exception to this 3-point rubric, see §6.1.)

4.2 A 4-Point Scale

The second refinement is to allow for four grades. I prefer to reserve a failing grade to indicate that the student did not do the work or that it was done with so little care that the work is the equivalent of stray marks on paper, not sufficient even for being called "wrong". This might include anything from an answer left blank to a partially, or even fully but randomly and completely incorrectly, filled-out truth table, depending on the instructor's preferences (as long as consistency is maintained).

So, the Triage Theory of Grading (now, perhaps, misleadingly called)[7] says that any item to be graded can best be graded on a 4-point scale:

Assignment done, and clearly adequate (full credit)	= 3
Assignment done, but not clearly adequate or inadequate (partial credit)	= 2
Assignment done, but clearly inadequate	= 1
Assignment not done	= 0

One of the earliest, if not the first, explicit university marking systems also used "four…and *only* four items", namely, "descriptive adjectives" used at Yale (c. 1785): (a) "*Optimi*" ("best", possibly in the sense of "best people" or "upper class"), (b) "second *Optimi*", (c) "*Inferiores*" ("inferior"), and (d) "*Pejores*" ("poorer, worse") (Smallwood 1935: 42–43; thanks to Spade for a translation suggestion). More recently, John Estell (n.d.) proposed a similar 4-point rubric for engineering education: 3 = "virtually no conceptual or procedural errors", 2 = "no significant conceptual errors and only minor procedural errors", 1 = "occasional conceptual errors and only minor procedural errors", 0 = "significant conceptual and/or procedural errors". Arthur Levine (1994) suggests a similar simplification: honors, high pass, pass, fail. But Levine's rubric is arguably guilty of grade inflation compared to my scheme, and Estell's rubric gives no credit where I would give minimal credit.

The crucial aspect of my theory of triage grading is that 0s, 1s, and 3s are intended to be clearly identifiable. Anything not clearly identifiable is a 2. Admittedly, there is some vagueness here: How "clear" must an answer be, in order to be, or not to be, "clearly adequate"?

If the student did not do the work (did not answer the question, did not even attempt to solve the problem, etc., or scribbled something incomprehensible or irrelevant on the answer sheet), that is clearly worth 0 points.

If the student did the work, but the answer is just plain wrong or shows no understanding of the issues, I would give it only 1 point. This is interpretable as giving the student 1 point for effort. (You could give it 0 points if you prefer *not* to distinguish an incorrect answer—which demonstrates that the student at least tried—from no answer at all.) There will typically be less vagueness about what counts as a "clearly wrong" or "clearly inadequate" answer than about what counts as a "clearly right" or "clearly adequate" answer.

If the student's answer is obviously adequate *or nearly so*, that should be worth the full 3 points. Here, I assume that, in many cases, there will be a clearly adequate answer. (I discuss ways to deal with essay questions in §4.3.) What does "*nearly* adequate" mean? This will depend on the nature of the question and the expected answer, but a good rule of thumb is this: If the student's answer, although not perfect, makes you think something along the lines of: "Yes, this student really seems to have a good, basic idea of what's going on with respect to this question", then it is nearly adequate and worth full credit.

If the student's answer is neither of the above, then—no matter how good or bad it is—give it 2 points. This is the only partial credit allowed. Two of the important advantages of the Triage Theory come from this: First, *you do not need to make fine distinctions among middling answers* or worry about whether a missing semicolon is important. This makes the evaluation and grading process much simpler. One potential problem is that some students (perhaps especially Perry Multiplists) might try to get partial (or even minimal) credit simply by writing down as much as possible, including, e.g., both correct and incorrect answers.[8] I am not sure that there is anything wrong with this. Writing down both a correct and an incorrect answer suggests that they *know* the answer even if they don't *realize that* they know it; arguably, that is worth partial credit. In any case, I doubt that any grading scheme can avoid this problem. With triage, we at least have a clear way to deal with it.

However, it is of the greatest importance that you—the grader—clearly indicate the errors that the student made and, perhaps, what a better answer would have looked like. Giving a student a less-than-perfect grade without indicating the errors is failing to do your job as a teacher. But you do not have to make fine distinctions among the incorrect answers.

The second such advantage is that, because each answer is worth only 3 points, where 3 = adequate and 2 = partial credit, and because no fractional points are allowed (the points are "quantum"—i.e., discrete—units), *a student cannot normally expect to get "just one more point" of partial credit.* A student who got only 1 point will either realize that the answer was so incorrect that partial credit is out of the question, or else will be able to persuade you that the answer was not, after all, so clearly incorrect that it was worth only 1 point. The latter case is the only one in which there is a possibility of your raising a grade, but the student must make a very good case, since, normally, there should be a very clear distinction between "clearly inadequate" and "partial credit". If the student got only 2 points, there is much less room to argue for full credit. Thus, one of the worst concomitants of grading—namely, dealing with unhappy students—is made much more bearable.

Why (these) 4 grades and not some other number? Large numbers simply make too many distinctions, ranging from the 101 points of a percentage scale (0%–100%), through the 20–40-point scales "to assess moral or staying power of a military unit" in historical, military-simulation, war games that was simplified to a 4-point system that was more efficient and equally meaningful,[9] to the 13 letter grades of the American system ('A', … , 'D' with + and –, and 'F').

A grading scheme of 5 grades (as the correspondent from the public-school system mentioned in §6.1, below, suggested)—in which 0 = assignment not done, 1 = assignment done but clearly incorrect, 2 = assignment done but *somewhat* incorrect, 3 = assignment done and *nearly* correct, and 4 = assignment done and clearly correct—makes too fine a distinction between the 2 ("somewhat") and 3 ("nearly") rankings. One of the strengths of the triage system (which would combine 2 and 3 into a single ranking) is that the only two decisions that a grader needs to make are "Is it clearly correct?" and "Is it clearly incorrect?". If the answer to both is "no", then it goes into the middle ranking. On the 5-point scheme, the grader needs to ask a further question—one with no clear answer—requiring distinguishing between "somewhat incorrect" and "mostly correct".

Walvoord & Anderson (1998) offer several other suggestions:

1. Use 6 grades: 'A', … , 'F', including 'E', *without* + or –.

   (But if 5 grades are too vague, surely so are 6.)

2. Use 4 grades: √, √+, √–, nil.

   (This seems akin to the triage system, but is not singled out by them for special treatment, and no interpretation is offered.)

3. Use 3 grades: "outstanding", "competent", "unacceptable".

   (These particular labels have undesirable ethical overtones.)

4. Use 2 grades: pass, fail.

   (This is fine in certain circumstances, but it is not very informative to the student.)

But they also point out that "The basic rule is to use the lowest number of grading levels consonant with your purpose and with student learning. It is easy to assume that, because at the end of the course you must assign grades in a thirteen-level system, every grade along the way must be calibrated on the same system" (p. 122). Of course, you *don't* have to give all 13 grades.

The triage system satisfies two of Walvoord & Anderson's desirability criteria (p. 72): It helps make grading "consistent and fair",[10] and it "saves time", thereby making grading efficient. Moreover, the triage theory is not "competitive grading [that] deemphasizes learning in favor of judging" (Krumboltz & Yeh 1997): The grades are (relatively) absolute (see §6.1) and convey reasonably precise information to the student about what they have learned and what they have not (yet) learned. Alfie Kohn (1994) praises "feedback" as a legitimate educational goal, while rejecting grading as a means to that end. Triage grading can serve this end in a positive fashion, however, as long as the students understand the meaning of the three levels, namely: You understand virtually all the material; you understand some, but not all, of the material; you understand virtually none of the material; you didn't even try. (See also §6.3.)

4.3 Assignments with Multiple Parts

The third refinement is to adapt triage grading to assignments with multiple parts. Suppose that you have given a homework problem set or an exam with 10 equally weighted questions. If each were worth 3 points on a 0,1,2,3-point scale as described above, then full credit would be 30 points. A student who got partial credit on each problem would get a score of 20 points. A student who tried all problems but got all of them wrong would get a score of 10 points. Only the students who did not do the work would get 0 points. (Of course, many students will get scores in between these: A student who failed to answer some questions but answered the others with only partial credit would get between 0 and 10 points, and so forth.) If you prefer to give percentages rather than raw scores, then 30 = 100%, 20 = 67%, 10 = 33%, etc. These numbers, whether percentages or raw scores, clearly measure, and hence indicate to the student, *how much of the assignment was done successfully.*

Should wrong answers be given more credit than no answer at all? The answer will depend in part on what behavior you are trying to encourage. I prefer to encourage my students to try to answer all problems, even if the attempt fails. Partial or completely wrong answers (such as those mentioned above in §4.2) are potentially more informative about a student's (lack of) knowledge than a blank page, though I will admit that an answer of, say, '42' to the question "What is the truth value of this proposition?" is probably best considered to be equivalent to no answer at all.

Similar techniques can be applied to other multiple-part assignments. For instance, a philosophy essay that consists of an analysis of an argument[11] might be graded as follows (though I will suggest some further qualifications below to make the grading of such an essay a bit more reasonable): Suppose that the argument has two premises and a conclusion. Here is one scheme for grading it (refinements should be obvious):[12]

Identification of premise 1	0,1,2,3
Evaluation of premise 1:	
Do you think that premise 1 is true? false?	0,1,2,3
Your reasons for your belief about its truth value	0,1,2,3
Evaluation of premise 2:	
Do you think that premise 2 is true? false?	0,1,2,3
Your reasons for your belief about its truth value	0,1,2,3
Evaluation of the conclusion:	
Does the conclusion follow validly from the premises?	0,1,2,3
Do you agree with the conclusion?	0,1,2,3
Your reasons for your agreement or disagreement	0,1,2,3
Total possible points = 24	

Another distinct advantage of this method is that it makes the grading of an essay such as this quite straightforward. Furthermore (and this, of course, is independent of my particular grading scheme), if the students are given such a grading rubric *before* they write the paper, they will have a much clearer idea of what is expected of them.

Perhaps, however, an essay assignment is more open ended, less argument-oriented. Still, you should have some idea of the kinds of things you will be looking for; each can be graded on the triage method. For example:

Originality of content	0,1,2,3
Organization	0,1,2,3
Spelling	0,1,2,3
Grammar	0,1,2,3

Total possible points = 12

Although such a breakdown may seem familiar, the key here is triage grading. You do not need to count the number of grammatical errors and select from a range of 13 or 101 grades. A simple decision—about whether the student has a generally good grasp of grammar (modulo a semicolon or two), a poor grasp, or somewhere in between—suffices. And the same goes for spelling and—perhaps more significantly—even for such vague areas as organization and originality: Any criterion is either clearly satisfied, clearly not satisfied, or is somewhere in between, the exact location in between being unimportant.[13]

4.4 Weights

One subtlety ignored in the previous analysis is that some parts of an assignment are usually more important than others. To reflect this, those parts can be weighted more heavily. Before indicating how this can be done on the Triage Theory, there is a possible objection to weighting to be considered.

Suppose, to make things simple, that you have a two-part assignment, with one part considerably harder or longer than the other. (This might not be a very well-considered assignment, but let's ignore that for now.) Should each part be worth the same number of points, or should the harder part be worth more? Or should it be worth *less*?

Multiplistic students would certainly want it to be worth more; after all, they will put more time into it, and such students believe that the grade on a problem should be directly proportional to the amount of time spent on it. And if it is really a harder problem than the other one, students should be amply rewarded for getting it right. On the other hand, students who get it wrong may be overly penalized.

To compensate for the imbalance between the two parts of the assignment, perhaps they should be weighted equally.[14] Granted, the student who gets the hard part right may not be rewarded as much as they deserve (perhaps more accurately: as much as they *feel* that they deserve), but this is balanced by not overly penalizing the students who did not get the hard part right. I normally favor equal weighting in cases such as this, and even Multiplistic students tend to agree that there is a certain amount of fairness in this (especially if they did not get full credit!).

But another approach is to split the hard problem up into smaller sub-problems, grading each on the 0–3-point scale. This has the effect of weighting the hard problem more, yet allows for finer distinctions of partial credit without giving up any of the advantages of the quantum aspect of the Triage Theory.

An alternative way to adapt the Triage Theory to differently weighted parts of an assignment is to multiply the points for that problem by some factor representing its relative weight. In the argument-analysis-essay example, above, if the instructor feels that reasons for believing (or doubting) the premises and evaluation of validity are far more important (say, 5 times more) than anything else, the instructor might use this:[15]

Identification of premise 1	0,1,2,3
Evaluation of premise 1:	
Do you think that premise 1 is true? false?	0,1,2,3
Your reasons for your belief about its truth value	0,5,10,15
Evaluation of premise 2:	
Do you think that premise 2 is true? false?	0,1,2,3
Your reasons for your belief about its truth value	0,5,10,15
Evaluation of the conclusion:	
Does the conclusion follow validly from the premises?	0,5,10,15
Do you agree with the conclusion?	0,1,2,3
Your reasons for your agreement or disagreement	0,5,10,15
Total possible points = 72	

It is important to remember, and to emphasize to the students, the quantum nature of these points. In the example above, not only is it impossible to get more than 2 but less than 3 points on the premise-identification part, *it is also impossible to get more than 10 but less than 15 points on the reasons part.* You simply explain to the students that their reasons were either clearly good (e.g., supportive and at least plausible), clearly bad (e.g., not supportive or clearly false), or somewhere in the vast in-between.

Another advantage of the Triage Theory is that it gives the student more information than some arbitrary number of points does: a 3 says “you got it right (for all practical purposes)”, a 2 says “almost, but not quite”, a 1 says “nope”, and a 0 says “you didn’t even try”; various weightings indicate relative importance. On a hypothetical 10-point scale, what is the significance of the difference between a score of 6 and a score of 7? As Thomas M. Haladyna (1999: 61) observes, “In other words, is the person who scores 74 on a high school writing graduation test and fails by one point really any different from the kid who scores a 75 and barely passes?"

What about an assignment (especially an essay) that gets a fairly low grade when graded more or less objectively as above but for which you, the instructor, feel it deserves something more or deserves some grade representing your overall impression? Nothing in the Triage Theory prevents you from including a completely subjective "fudge factor" and assigning it 0, 1, 2, or 3 points (perhaps weighted), as long as the "fudge factor" is taken into account for all students on that assignment. As Jonathan Bona pointed out to me, this can also be used to raise everyone's grade if the instructor feels that the assignment was harder than expected. (This *may* be a rare legitimate use of "curving"; cf. §6.2.) On the other hand, use of a fudge factor runs the risk of "students…pressing the grader to increase their fudge points" (Bona, personal communication, 2008). Thus, it should be used sparingly, if at all.

5 The Triage Theory: Letter Grades

So much for numerical points. How do I convert this to letter grades (which my university requires)? Here's my principle, which is independent of the above point-grading scheme and makes several arbitrary assumptions. Since:

3 = assignment done, and clearly adequate

2 = assignment done, but neither clearly adequate nor clearly inadequate

1 = assignment done, but clearly inadequate

0 = assignment not done

I take:

3 = A

2 = C

1 = D

0 = F

The mappings to 'F', 'D', and 'A' may be obvious, but what happened to 'B'? We could map the 4-number scheme into the 5-letter scheme as follows:

3 = A

$2\frac{1}{3}$ = B

$1\frac{2}{3}$ = C

1 = D

0 = F

But then should 2 = 'C+' or should 2 = 'B–'? And dealing with fractions violates the quantum principle. (A similar mapping according to which 'A' = 3, 'B' = $2\frac{1}{4}$, 'C' = $1\frac{1}{2}$, 'D' = $\frac{3}{4}$, and 'F' = 0 seems worse.)

Why should 2 map to 'C' rather than 'B'? Because 'C' is supposed to be "average". Here, 'average' does not necessarily mean the arithmetic mean. Instead, I intend it (as does my university's undergraduate catalog; see §6.1) in the sense of "usual", "ordinary", "intermediate".[16] Indeed, an 1842 "account" of marks at Yale states: "marks range from 0 to 4. 2 *is considered as* the average; and a student not receiving this average…is obliged to leave…" (quoted in

Smallwood 1935: 47, my italics). 2 is, indeed, intermediate between the extremes of "adequate" (= 3) and "inadequate" (= 1). (Of course, given the numerical grading scheme (and not counting 0), it is also the arithmetic mean.)

(If you wish to include 0, then perhaps 'C' should be 1.5. But that introduces fractions or decimals, which makes for a certain awkwardness and suggests a greater level of precision in the grading scheme than there really is. Alternatively, one could insist that 1.5 *is* average, and then define a 2-point 'C' as "slightly above average".)

If 'A' = 3 and 'C' = 2, and if point-assignments are quantized (i.e., no fractional points), then how does a student get a grade of 'B'? If enough assignments during a semester are given using this letter-grade scheme, 'B' grades will appear when grades over several assignments get averaged (in the arithmetic sense). They will also appear, as will + and – grades, if the total number of points for a given assignment is large enough, using the mapping described below. (A word of warning: Things get a bit technical, somewhat arbitrary, and even slightly inelegant at this point. Those who do not intend to adopt triage grading can skip the details.)

Usually, each assignment is worth a multiple of 3 points (e.g., if there are 10 parts worth 3 points each, then the total = 30 points). Sometimes a question is of the "true-false" variety, where there is no opportunity for partial credit. I will not consider here whether this is a good idea or not; sometimes it seems appropriate or unavoidable. Such questions can be graded as either 0, 1, or 3 (i.e., no answer, incorrect, or correct), with no possibility of 2 points (no partial credit). But sometimes a weighting scheme or a very simple problem suggests an even simpler point assignment of, say, 0 (i.e., incorrect) or else 1 (i.e., correct).

To map the 0,1,2,3-point scale into letter grades, let n = the number of parts (in our example, $n = 10$), and let T = the total number of points (so, $T = 3n$; in the example, $T = 30 = 3*10$). Then $3n$ maps to 'A', $2n$ maps to 'C', n maps to 'D', and 0 maps to 'F'. Other grades can be interpolated in an evenly spaced fashion, as shown in Table 1.

PUT TABLE 1 HERE

This table represents the mapping that I use from the numerical scheme to the letter scheme that is used at my university, where there are no grades of 'A+' or 'D–'; the table is explained below. Other interpolation schemes may be necessary for other letter grades, and, indeed, other interpolation schemes are possible even for the letters shown below. (There is a certain amount of unavoidable subjectivity in any aspect of grading; more on this in §6.1, below.) Incidentally, triage effectively eliminates most 'F's except as a message that the student did no work (but see below, this section).

Table 1 needs a bit of explanation. The first column, "Factor", is based on n, the multiple of 3 that is such that $3n$ = the total score. Because a raw score of $3n$ is clearly full credit, it is mapped to a grade of 'A' (shown in the second column, "Grade"). Similarly, $2n$ is mapped to 'C', n to 'D', and 0 to 'F', in accordance with my analysis above. (This, of course, is an arbitrary and subjective mapping; it is up to you to choose the factor-to-grade mapping.)

The next question is how to interpolate the other letter grades. I assume that 'B' should be halfway between 'A' and 'C'; thus, it corresponds to a "factor" of $\frac{5}{2}n$. You could, of course, make a different assumption about where 'B' should be interpolated. But given that 2 is halfway between 3 and 1, I see no reason not to treat other intermediate grades in the same equidistant fashion. With 'B' halfway between 'A' and 'C', I similarly interpolate 'A–' and 'B+' equally spaced between 'A' and 'B'; this results in a raw score of $\frac{8}{3}n$ (= $\frac{16}{6}n$) being mapped to 'B+' and $\frac{17}{6}n$ being mapped to 'A–'.[17] (Again, these letter grades need not be mapped equidistantly; I merely choose to do so.) Similarly, if 'C+' and 'B–' are interpolated equidistantly between 'C' and 'B', they map to raw scores of $\frac{13}{6}n$ and $\frac{7}{3}n$ (= $\frac{14}{6}n$), respectively. Finally, mapping 'D+' and 'C–' equidistantly between 'D' and 'C' maps them to raw scores of $\frac{4}{3}n$ and $\frac{5}{3}n$, respectively. This completes the explanation of the first two columns.

The third column, "Point Range" is the most useful for actually assigning letter grades based on raw scores. Again, however, I have made certain assumptions that others might make differently. Here, the question is how to map *from* raw scores that are intermediate between the ones identified above *to* letters. The problem is that, although there might be raw scores of (say) 28 or 29 on an assignment whose points total 30, it is not immediately obvious whether they should be mapped to ‘B+’, ‘A–’, or ‘A’. This problem is not unique to triage grading; if a student has a 3.5 GPA (or QPA), should that be considered an ‘A–’ (= 3.7) or a ‘B+’ (= 3.3)?[18]

For my analysis here, I found it easier to think in terms of T, the total score (recall that $T = 3n$). Consider the interval between ‘A–’ and ‘A’. The raw-score endpoints are $\frac{17}{18}T$ and T. Because T is clearly an ‘A’, the question is whether $\frac{17}{18}T$ should be the lowest ‘A’ or the highest ‘A–’. (This is the same issue discussed in §4.1, namely, should the intervals be closed at the “high” end or the “low” end?) In the interests of curbing grade inflation, however small, I chose to make the low endpoint a high ‘A–’. Because the raw scores are integers, the lowest ‘A’ is therefore a raw score of $\frac{17}{18}T + 1$, the lowest ‘A–’ is $\frac{8}{9}T + 1$ (= $\frac{16}{18}T + 1$), etc.[19] This works till we get down to ‘D’, whose high raw score must (based on my assumptions) be $\frac{1}{3}T$. I then assume that the range between 0 and $\frac{1}{3}T$ is more-or-less evenly split between ‘D’ and ‘F’; thus, ‘D’ ranges from $\frac{1}{6}T + 1$ to $\frac{1}{3}T$ (= $\frac{2}{6}T$), and ‘F’ ranges from 0 (which has to be its low endpoint) to $\frac{1}{6}T$. (This perhaps violates my principle that ‘F’ should be reserved for “no work” and ‘D’ for “some work, but clearly inadequate”. So it goes; you may decide otherwise.)

I use the “Point Range” column for grading: Given T, I create a chart showing the range of raw scores and their corresponding letter grades. For example, to take the 72-point argument-analysis essay from §4.4, above, I would use the letter-grade equivalents shown in Table 2 (T = 72).

PUT TABLE 2 HERE

There can be no “borderline” scores that could map to more than one letter grade. However, there are often cases where a student gets the highest score for a given letter but cannot be given “just one more point” to be pushed over to the next highest letter grade. This happens when the only way to get that one extra point would be to change a “clearly wrong” grade on some problem to a “partial credit” grade, or a “partial credit” grade to a “clearly right” grade. And the whole Triage Theory has been designed to make that difficult, if not impossible. When I explain this to (unhappy) students, they usually understand, because the grading system is clear and fair.

That said, I should also say that I occasionally promise that, at the end of the semester, when all the grades are in and I am computing the student’s final course grade, if that one point would have made the difference between one final letter grade and the next highest one (say, between an ‘A–’ and an ‘A’), I will give the student the higher grade. This almost never happens; when it does, it seems to me to be a perfectly reasonable thing to do. (This perhaps violates my desire to curb grade inflation. As I said before, so it goes; you may decide otherwise.) Alternatively, of course, one can use some non-graded achievement (e.g., attendance or class participation) to raise a borderline grade (as long as it is done consistently for all students).

The fourth column, “T = 100%”, shows the mapping of *percentages* to letter grades (i.e., when $T = 100$). Here, it can be seen that ‘A’ maps to the highest 5%, ‘B’ to the low 80s, ‘C’ to the high 50s–middle 60s, and ‘D’ to the 20s (with a bit of overflow into both the high teens and low 30s). Two things are apparent: This is not a “normal” (or “curved”) distribution (see §6.2), nor is it the “classical” mapping rejected above. As I said in §4.1, I have never understood the classical mapping; it seems completely arbitrary. The Triage Theory at least has a rational basis.

The final column, “Width”, represents the number of different possible points that correspond to each letter grade. This only emphasizes the difference between the Triage Theory and “normal” distributions. In some sense, it is easier to get an ‘F’ than it is to get an ‘A’: $\frac{1}{18}T$ is the distance between the ‘A–’ and ‘A’ endpoints, and similarly for the other “widths” (with some adjustments for rounding, as noted in the caption for Table 2).

Because I only use the “grade” and “range” columns to map point-values to letters, as in Table 2, I never need to fill in the complete Table 1 in practice. However, for the sake of clarity,

Table 3 is an instance of Table 1 with all values filled in for the 72-point assignment (here, $T = 72$, $n = \frac{1}{3}T = 24$).

PUT TABLE 3 HERE

6 Other Issues in Grading

6.1 On the Subjectivity of Grades

Recall Robert Paul Wolff's "three species of grading", introduced in §2. The second is "evaluation", that is, "the measuring of a product or performance against an independent *objective* standard of excellence" (Wolff 1969: 59, my emphasis). However, "evaluation…is external to education properly so-called" (Wolff 1969: 64); that is, assigning a symbol to the critiques adds nothing to the critique, an observation reminiscent of emotivism in ethics: 'Good' is just a positive utterance; if you do good work, as determined by "criticism", then *calling* it 'good' adds no information. But we teachers are *forced* to summarize our educational critiques. Hence, grading as evaluation is inevitable in *our* society, as opposed to the revolutionary one that Wolff wants. Clearly, triage is grading as evaluation, but I dispute the objectivity of the standard (in most cases).

Grades would be objective if there were some absolute scale on which students were graded, or if all (or some significant number) of graders would independently agree about a student's grade. (This would be closer, perhaps, to what Kant called "intersubjectivity", or what social scientists call "inter-rater reliability".) But there is no absolute scale. All grading is relative, hence subjective.

However, the academic institution where you teach will have a culture and a set of grading expectations that you might not share. For instance, one correspondent in a pubic-school system in the southeastern US told me that his school requires all faculty to use the following percentage-to-letter mapping: <70% = 'F', 70–76% = 'D', 77–84% = 'C', 85–92% = 'B', and

93–100% = 'A'. Not only can such institutional-vs.-individual differences be "ethically troubling",[20] but different schools with different percentage-to-letter mappings will not have comparable grades.[21] I have the luxury of being able to assign my own letter grades however I see fit, the only requirement being that they conform to my university's English-language interpretation ('A' = "high distinction", 'C' = "average", etc.).[22] But those English words are vague, so I have some freedom. My correspondent, on the other hand, has to abide by those percentage equivalents to letters. It seems to me that there are three (bad) options:

1. Use the school's percentages instead of more stringent ones that the instructor might prefer to give; in this case, that would give the students slightly higher grades than the instructor's scheme would give them.
2. Use the instructor's own scheme, which would give the students lower letter grades than they might otherwise get.
3. Give the students two sets of grades: An "internal" grade using the instructor's scheme, which gives the students hopefully useful feedback on how the instructor thinks they are really doing, and an "external" grade using the school's scheme, for official purposes. This amounts to "curving" the instructor's grades.

One piece of advice is to make your standards clear at the outset and to explain to the students what you are measuring. For me, an 'A' (or 3 points) represents complete or nearly complete understanding or mastery of the subject, a 'D' (or 1 point) represents some effort but little or no success in understanding, and a 'C' (or 2 points) represents everything in between. (An 'F', or 0 points, represents complete, or nearly complete, lack, of effort.)

Another piece of advice is to stand firm in your belief that you have the qualifications to make this kind of judgment. Such subjectivity is not inherently evil. Walvoord & Anderson (1998: 11) observe that teachers must "substitute judgment for objectivity". That is, because all grades are subjective, and the teacher is an "informed professional", it is the *teacher's* judgment about what the student has learned that is the measuring stick. The grades are relative to *your* students, and you have a right to those standards because, and to the extent that, you are an informed professional.

Here, issues of reliability and validity enter: Ideally, my grading judgments should match those of other equally qualified instructors, and they should be reasonably consistent over time. On the triage scheme, differences should be no greater than 1 point (e.g., two instructors might disagree over full vs. partial credit, but should not disagree over full vs. minimal credit). But "professional judgment" is more a matter of *assessment* (which is necessary for learning) than of reliability vs. validity.

The trick is to minimize the subjectivity. The Triage Theory is an attempt to do this by limiting the choices that a grader has to make and by preventing (or relieving) the grader from having to make fine distinctions within the vast category of "neither clearly adequate nor clearly inadequate". Still, there are some subjective calls to make, and you might very well disagree with the way that I have made them. For instance, you might not accept my (subjective) mapping from numerical points to letters. You might not accept my schemes for weighting parts of problems or for dealing with easy vs. hard problems. But none of these choices are essential parts of the Triage Theory.

6.2 Grading on a Curve

Although it is not directly relevant to the Triage Theory, I want to say a few words about "grading on a curve". As I understand this practice, it makes a student's grade relative, not to some external or instructor-based standard, but to the other grades in the course.

The idea behind grading on a curve is that the course grades should be distributed along a bell curve: Most of the grades should be 'C', a smaller—but relatively equal—number should be 'B' or 'D' (i.e., there should be roughly the same number of 'B's as 'D's), and a very few (but relatively equal in numbers) should be 'F' or 'A'. The best students will get 'A', the average ones will get 'C', and the worst will fail.

This, it seems to me, gives the student little information that is of any use. If all the other students are worse than you, then you will do well, even if you did poorly on any "objective" scale; and, if all the other students are better than you, then you will do poorly, no matter how much you learned or how smart you were in the course. (See also Kohn 1994.)

Grading on a curve is a kind of ranking: it is Wolff's third "species of grading": "a relative comparison of the performances of a number of students, for the purpose of determining a linear ordering of comparative excellence" (Wolff 1969: 61–62; cf. Kohn's (1994) criticism of grading as "sorting"). Ranking is inevitable once there is "evaluation". It "performs a function which is neither professional nor educational, but merely…economic. …[It] facilitate[s] the fair allocation of scarce resources and utilities" (Wolff 1969: 65–66).

But if we stick with triage, we only have three ranks to be concerned with, which seems easier and more useful than the slippery slope leading to 13 varieties of letters or 101 varieties of percentages. There will still, of course, be large matters of "taste".

Here is a related myth: "*Since everyone cannot receive the same grade,* Ms. Smith [an 11th-grade English teacher] must find reasons to give some papers lower grades than others" (Krumboltz & Yeh 1997; my emphasis). Why make this assumption? Why force yourself to "look for flaws…[and] concentrate on the negative" (Krumboltz & Yeh 1997)? If everyone does equally good (or average, or bad) work, then everyone deserves equally full (or partial, or no) credit.

6.3 What Should We Tell the Students?

Partly because this grading scheme is rather different from what most students have seen, but also because I believe that students have a right to understand their instructors' grading schemes, I explain the triage theory briefly, and I publicize (on my syllabi) a website that outlines it.[23] I encourage discussion of it in my classes. (The encouragement is usually unsuccessful, which may actually indicate student satisfaction; otherwise, they would complain loudly.) Most students do not understand it at first, but they begin to see how it operates after their first graded assignment. I also provide grading rubrics to accompany each assignment. This not only lets the students know ahead of time how they will be graded, but it often gives them an outline of how to do the assignment; at least, it tells them what I am looking for.

7 Summary

The essence of the Triage Theory of grading is that an item to be graded can, and should, be graded only as either clearly adequate, clearly inadequate, or neither clearly adequate nor clearly inadequate, without making any finer distinctions.

Haladyna (1999: ix) says that "Before we assign a grade to any students, we need:

1. an idea about what a grade means,
2. an understanding of the purposes of grading,
3. a set of personal beliefs and proven principles that we will use in teaching and grading,
4. a set of criteria on which the grade is based, and, finally,
5. a grading method, which is a set of procedures that we consistently follow in arriving at each student's grade."

On the Triage Theory,

1. a grade measures how much the student has learned or understood, using a simple, 3-point scale,
2. the purpose of grading is to give that feedback to the student,
3. grading should not (indeed, cannot) be overly precise (cf. Postman), and should be understandable by the student.
4. The criteria are very simple: Has the student understood the material (not necessarily perfectly, but sufficiently well)? Or has the student completely failed to understand it? Or is the student somewhere in between these extremes?[24]

5. Finally, the grading method is to break complex assignments into simpler parts, each of which is (recursively) graded by triage.

8 Closing Remark

I close with a quote from a “Walnut Cove” comic strip of several years ago:

> Student, to teacher: “Can’t we curve this F up to a D?”
>
> Teacher, to student: “I don’t think you understand my responsibility as your high school teacher. Right now you are but a tottyheaded young lad. But someday you will be old enough to participate in society. Someday you may even run for president! That is where my duty as a conscientious educator comes in. It’s my job to stop you.”

Acknowledgments

I am grateful to Carl Alphonce, Jonathan Bona, Tanya Christ, Stephen Colbert (no, not that one!), Paul V. Gestwicki, Albert Goldfain, Timothy Grove, Stuart C. Shapiro, Thomas J. Shuell, Paul Vincent Spade, Karen M. Wieland, and an anonymous referee for comments on earlier versions of this essay; to numerous former teaching assistants who have welcomed the scheme, found it useful, and even adopted it; and to even more numerous former students who have been graded under it.

Endnotes

[1] For an explanation of the subtitle, see note 7.

[2] The use of capital letters in discussions of Perry's theory indicates terms as they are understood from the point of view of the students. It is analogous to the use of "scare quotes" (such as the ones I just used). For more on Perry's theory, see Rapaport 1982, 1984ab, 1987, and the references cited at [http://www.cse.buffalo.edu/~rapaport/perry-positions.html] (accessed 11 April 2011).

[3] As an anonymous referee pointed out.

[4] Spade, however, claims to no longer remember this (personal communication, 2008).

[5] Haladyna 1999 contains a good summary of pros and cons of the purposes, techniques, and varieties of grading.

[6] For an enlightening history, see Smallwood 1935, Ch. III, usefully summarized in Durm 1993.

[7] I had always thought that 'triage' meant to sort into *three* categories, but, as my colleague Carl Alphonce pointed out to me, it doesn't: According to the *Oxford English Dictionary Online* [http://www.oed.com/view/Entry/205658] (accessed 11 April 2011), 'triage' is not etymologically or semantically related to 'tri-' (meaning "three"), but comes from the French *trier* (meaning "to pick, cull"), which, in turn, is related to 'try' (in the sense "to sift or pick out"). Nevertheless, the *OED*'s earliest citation is a brief except from the 1728 Chambers *Cyclopaedia* entry on "Wool", which reads as follows:

> Each Fleece consists of Wool of divers Qualities and Degrees of Fineness, which the Dealers therein take care to separate. …If the Triage or Separation be well made, in fifteen Bales there will be twelve mark'd R, that is, Refine or Prime.

The ellipsis is in the *OED* citation; the full passage is on p. 377 of the online edition of the Chambers *Cyclopaedia* at [http://tinyurl.com/ChambersWool] (accessed 11 April 2011; or link to [http://digital.library.wisc.edu/1711.dl/HistSciTech.cyclopaedia02] (accessed 11 April 2011), then select "Weeping–wythe" and go to p. 377). The full passage clearly indicates that the triage is not only a "separation", but a sorting—indeed, a grading—into *three* categories:

> The *Spaniards* make the like division into three *Sorts*, which they call Prime, Second, and Third; and for the greater Ease, denote each Bale or Pack with a Capital Letter denoting the Sort—If the Triage or Separation be well made, in fifteen Bales there will be twelve mark'd R, that is Refine or Prime; two mark'd F, for Fine or Second; and one S, for Thirds. (Chambers *Cyclopaedia*: 377.)

My subtitle comes from the *OED*'s next citation, an 1825 issue of *Gentlemen's Magazine*, which also describes a tripartite triage:

> These [pickers] sort the [Coffee] berries into three classes; 'best quality', 'middling', and the third of all the bad broken berries..is called 'triage coffee'. (The 2-dot ellipsis is in the *OED* citation.)

[8] Albert Goldfain, personal communication, 2008.

[9] Timothy Grove, personal communication, 2009.

[10] Other things that contribute to overall grading fairness include "the characteristic being assessed and the weighing of the measurement in constructing a final course grade", as an anonymous reviewer pointed out.

[11]This is applicable to other disciplines, too: Computing Curricula 2001's "Social and Professional Issues" knowledge area includes the item "Methods and Tools of Analysis" (SP3), which covers argument-analysis techniques [http://www.acm.org/education/curric_vols/cc2001.pdf], accessed 6 April 2011.

[12] Programming projects in computer science that require a problem definition, a top-down design, documented code, and annotated output can also be graded this way:

Problem definition	0,1,2,3
Top-down design	0,1,2,3
Documented code:	
Code	0,1,2,3
Documentation	0,1,2,3
Annotated output:	
Output	0,1,2,3
Annotations	0,1,2,3
Total possible points = 18	

See note 15 for continuation of this example.

[13] Griffin 1998 has a similar 3-part rubric for student essays: highly successful, moderately successful, and less-than-successful. This seems close to my triage scale, as well as to Spade's original insight. (Thanks to Karen M. Wieland for pointing this out.)

[14] As my former math-methods professor, Anne Peskin, advocated.

[15] In the programming-project example from note 12, if the instructor feels that documented code is 5 times more important than anything else, the instructor might use this:

Problem definition	0,1,2,3
Top-down design	0,1,2,3
Documented code:	
Code	0,5,10,15
Documentation	0,5,10,15
Annotated output:	
Output	0,1,2,3
Annotations	0,1,2,3
Total possible points = 42	

[16] See the *Oxford English Dictionary* online entry for the adjective ‘average’, sense 2a [http://www.oed.com/view/Entry/13683] (accessed 11 April 2011), which includes this: “medium, ordinary; of the usual or prevalent standard”, or consider the idiomatic expression “average Joe” (cited in the same entry).

[17] The “distance” between the ‘A’ and ‘B’ endpoints is $3n - \frac{5}{2}n = \frac{1}{2}n$, so 1/3 of the way from ‘B’ to ‘A’, which is ‘B+’, would be $\frac{5}{2}n + \frac{1}{6}n = \frac{8}{3}n$. Similarly, 1/3 of the way from there to ‘A’, which is ‘A–’, would be $\frac{8}{3}n + \frac{1}{6}n = \frac{17}{6}n$.

[18] For that matter, should ‘B+’ be 3.33 or 3.34 instead? If so, then perhaps a 3.50 GPA should be ‘B+’. But then what about a 3.51 or 3.52 GPA?

[19] E.g., if $T = 90$, then a raw score of $\frac{17}{18}*90 = 85$, so the lowest ‘A’ would be 86, the highest ‘A–’ would be 85, the lowest ‘A–’ would be $\frac{8}{9}*90 + 1 = 81$, etc.

[20] Tanya Christ, personal communication, 2008.

---

[21] An anonymous referee argues that:

> assuming that the scales are linear, that they have fixed end-points, and that the intervals above failing on a given scale are equal in size, the different schools will certainly have comparable grades. One instructor might grade on a "33.33" pass like the author, I grade on a 50-pass, and my colleague down the hall grades on a 70-pass, but our grades are all easily converted from one scale to another, arithmetically no different than converting Celsius to Fahrenheit.

I doubt that all graders use equal-sized intervals above failing; surely, those who grade on a curve don't. Moreover, the real issue is epistemological, not metaphysical. I stand by my statement in the text, but am willing to modify it: Different schools with different mappings may *have* comparable grades, but there may be no way to know how to compare them.

[22] [http://undergrad-catalog.buffalo.edu/policies/grading/explanation.shtml], accessed 4 April 2011.

[23] [http://www.cse.buffalo.edu/~rapaport/howigrade.html], accessed 4 April 2011.

[24] It can also provide a guide to the instructor for designing questions that differentiate well between nearly full understanding, failure to understand, and partial understanding (Goldfain, personal communication, 2008).

References

Batiuk, Tom (2004, March 9), "Funky Winkerbean" comic strip.

Clio, Max (2004, June 25), "On Your Marks", *Chronicle of Higher Education*. See also:

Clio, Max (2003, November 21), "Grading on My Nerves", *Chronicle of Higher*

Education.

Cohen, Randy (2005, April 3), "The Art of Grading", *The New York Times Magazine.*

Durm, Mark W. (1993), "An A Is Not an A Is Not an A: A History of Grading", *Educational Forum* 57 (Spring).

Estell, John K. (n.d.), "Streamlining the Assessment Process with the Faculty Course Assessment Report" [http://www.engr.ncsu.edu/abetcommittee/resources/97_Estell_John_Materials%20fcar.pdf] (accessed 11 April 2011).

Griffin, C. Williams (1998), "Improving Students' Writing Strategies: Knowing versus Doing", *College Teaching* 46(2) (Spring): 48–52.

Haladyna, Thomas M. (1999), *A Complete guide to Student Grading* (Boston: Allyn & Bacon).

Hargis, Charles H. (1990), *Grades and Grading Practices: Obstacles to Improving Education and to Helping At-Risk Students* (Springfield, IL: Charles C Thomas).

Kohn, Alfie (1994), "Grading: The Issue Is Not How but Why", *Educational Leadership* 52(2) (October): 38–41.

Krumboltz, John D.; & Yeh, Christine J. (1996), "Competitive Grading Sabotages Good Teaching", *Phi Delta Kappan* 78(4): 324–326.

Lakoff, George (1987), *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind* (Chicago: University of Chicago Press).

Levine, Arthur (1994), "To Deflate Grade Inflation, Simplify the System", *Chronicle of Higher Education* (19 January): B3.

Mervis, Carolyn B.; & Rosch, Eleanor (1981), "Categorization of Natural Objects", *Annual Review of Psychology* 32: 89–115.

Perry, William G. (1970), *Forms of Intellectual and Ethical Development in the College Years: A Scheme* (New York: Holt, Rinehart, and Winston).

Perry, William G. (1981), "Cognitive and Ethical Growth: The Making of Meaning", in A. W. Chickering (ed.), *The Modern American College* (San Francisco: Jossey-Bass): 76–116.

Postman, Neil (1992), *Technopoly: The Surrender of Culture to Technology* (New York: Vintage Books).

Rapaport, William J. (1982), "Unsolvable Problems and Philosophical Progress", *American Philosophical Quarterly* 19: 289–298.

Rapaport, William J. (1984a), "Can Philosophy Solve Its Own Problems?", *The [SUNY] News* 13 (May/June): F2–F3 [http://www.cse.buffalo.edu/~rapaport/Papers/rapaport84b.canphilsolve.pdf], accessed 11 April 2011.

Rapaport, William J. (1984b), "Critical Thinking and Cognitive Development", *Proceedings of the American Philosophical Association* 57: 610–615.

Rapaport, William J. (1987), "Philosophy for Children and Other People", *American Philosophical Association Newsletter on Teaching Philosophy* (Summer): 19–22 [http://www.cse.buffalo.edu/~rapaport/Papers/phil4children.pdf], accessed 11 April 2011.

Smallwood, Mary Lovett (1935), *An Historical Study of Examinations and Grading Systems in Early American Universities: A Critical Study of the Original Records of Harvard, William and Mary, Yale, Mount Holyoke, and Michigan from Their Founding to 1900* (Cambridge, MA: Harvard University Press).

Walvoord, Barbara E.; & Anderson, Virginia Johnson (1998), *Effective Grading: A Tool for Learning and Assessment* (San Francisco: Jossey-Bass).

Wolff, Robert Paul (1969), "A Discourse on Grading", in Robert Paul Wolff, *The Ideal of*

the University (Boston: Beacon Press), Part Two, Ch. One, pp. 58–68.

Factor	**Grade**	**Point Range**	***T* = 100%**	**Width**
$3n$ (= T)	A	from ($\frac{17}{18}T + 1$) to T	95–100	$\frac{1}{18}T$ (= $\frac{1}{6}n$)
$\frac{17}{6}n$ (= $\frac{17}{18}T$)	A–	from ($\frac{8}{9}T + 1$) to $\frac{17}{18}T$	90–94	$\frac{1}{18}T$ (= $\frac{1}{6}n$)
$\frac{8}{3}n$ (= $\frac{8}{9}T$)	B+	from ($\frac{5}{6}T + 1$) to $\frac{8}{9}T$	84–89	$\frac{1}{18}T$ (= $\frac{1}{6}n$)
$\frac{5}{2}n$ (= $\frac{5}{6}T$)	B	from ($\frac{7}{9}T + 1$) to $\frac{5}{6}T$	79–83	$\frac{1}{18}T$ (= $\frac{1}{6}n$)
$\frac{7}{3}n$ (= $\frac{7}{9}T$)	B–	from ($\frac{13}{18}T + 1$) to $\frac{7}{9}T$	73–78	$\frac{1}{18}T$ (= $\frac{1}{6}n$)
$\frac{13}{6}n$ (= $\frac{13}{18}T$)	C+	from ($\frac{2}{3}T + 1$) to $\frac{13}{18}T$	68–72	$\frac{1}{18}T$ (= $\frac{1}{6}n$)
$2n$ (= $\frac{2}{3}T$)	C	from ($\frac{5}{9}T + 1$) to $\frac{2}{3}T$	57–67	$\frac{1}{9}T$ (= $\frac{1}{3}n$)
$\frac{5}{3}n$ (= $\frac{5}{9}T$)	C–	from ($\frac{4}{9}T + 1$) to $\frac{5}{9}T$	45–56	$\frac{1}{9}T$ (= $\frac{1}{3}n$)
$\frac{4}{3}n$ (= $\frac{4}{9}T$)	D+	from ($\frac{1}{3}T + 1$) to $\frac{4}{9}T$	34–44	$\frac{1}{9}T$ (= $\frac{1}{3}n$)
n (= $\frac{1}{3}T$)	D	from ($\frac{1}{6}T + 1$) to $\frac{1}{3}T$	18–33	$\frac{1}{6}T$ (= $\frac{1}{2}n$)
0	F	from 0 to $\frac{1}{6}T$	0–17	$\frac{1}{6}T + 1$

Table 1: From *T* Points to Letters

(Note that a grade of 'A' would be given to any numerical grade in the range from ($\frac{17}{18}T + 1$) to T, where T = the total number of points on the assignment. For an example where T = 72, see Table 2.)

Grade	Range
A	69–72
A–	65–68
B+	61–64
B	57–60
B–	53–56
C+	49–52
C	41–48
C–	33–40
D+	25–32
D	13–24
F	0–12

Table 2: Letter-Grade Equivalents When *T* = 72.

(Point-values can be rounded to the nearest whole number. E.g., the low end for 'D+' can be rounded up, but the low end for 'C–' might be rounded down. Alternatively, one could be strict (rounding all such fractional point-values up) or lenient (rounding them down).)

Factor	Grade	Range	T = 100%	Width
72	A	69–72	95–100%	4 (i.e., 4 grades will get ‘A’: 69, 70, 71, 72)
68	A–	65–68	90–94%	4
64	B+	61–64	84–89%	4
60	B	57–60	79–83%	4
56	B–	53–56	73–78%	4
52	C+	49–52	68–72%	4
48	C	41–48	57–67%	8
40	C–	33–40	45–56%	8
32	D+	25–32	34–44%	8
24	D	13–24	18–33%	12
0	F	0–12	0–17%	13

Table 3: From 72 Points to Letters.

(= Table 1, where $n = 24$, $T = 72$)