A Triage Theory of Grading: The Good, the Bad, and the Middling*

William J. Rapaport

**Department of Computer Science and Engineering,
Department of Philosophy, Department of Linguistics,
and Center for Cognitive Science
State University of New York at Buffalo, Buffalo, NY 14260-2000**

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

March 23, 2009

Abstract

This essay presents and defends a triage theory of grading: Any item to be graded should get full credit if it is clearly or substantially correct, minimal credit 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. Other issues discussed include grading on a curve, the subjectivity of grading, and reasons for and against grading.

*For an explanation of the subtitle, see note 7.

1 Problems with Grading

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

> I'm planning to outsource my class load next year. I'm thinking India for paper correction. (Burke 2007.)

> "Mr. Moore… 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.)

I hate to grade.

In this essay, I present a "universal" grading technique (i.e., one applicable to any discipline), inspired by a casual remark made by one of my former professors, that 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'). These are "quantum" units; there are no grades in between these.

Why do we grade? Because "[w]e have a powerful need to grade" (Hargis 1980: 3). Perhaps it is just an aspect of the human propensity to classify or categorize (or is it a cognitive imperative, perhaps 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" cartoon, 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 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.)

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 in such a school: 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[1] to questions posed by Authorities (i.e., by us teachers). Authorities are seen by such students as teaching

[1] 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 "square quotes" (such as the ones I just used).

by giving them the answers; the students see their job as repeating the answers when asked for them. If they repeat The Correct Answer, 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).

But Dualistic students eventually come to see that there are grey areas, that there are questions whose answers we don't know *yet*: They take the position of "Multiplism". 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.[2]

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 §5.) 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 §5) 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 missing a semicolon? (This is a classic conundrum for teachers of computer programming, especially because such errors can be found—and automatically corrected—by computers.) 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−')?

The Triage Theory of Grading resolves most of these issues.

[2]Multiplists are further discussed below in notes 9 and 10. I leave for another time the question of how students at other Perry positions (especially "Contextual Relativism"—the position taken by those students who have come to see that all claims must be understood relative to, and in the context of, the evidence that supports them) might view grading.

2 The Triage Theory of Grading: Origin and Outline

The Triage Theory 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.[3] 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'.

Thus, the fundamental insight is to give only three grades, and no grades in between. Why three and not, say, two? "It 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. (p. 60.)

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 among the bottom, *and also no clear differences within the middle*. "[N]o standard, whether pass/fail or letter grades, makes a real delineation

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

between groups of students" (Haladyna 1999: 61).[4] 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.

3 The Triage Theory: Details

3.1 Numerical Grading

The first refinement is to grade numerically, not by percentages and not by letters (at least not initially; see §4). 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.[5]

I do not wish to make this assumption, because I see no rationale for it. This "classical" mapping of percentages to letters is probably a recent invention. (For an enlightening history, see Smallwood 1935, Ch. III, usefully summarized in Durm 1993.) 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: 100 ("perfect"), 75–99, 51–74, 26–50, and 25 or below (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 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:

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

[5] 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. I.e., 100% $\leq$ 'A' $\leq$ 90%, and 90% $<$ 'B' $\leq$ 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; e.g., 0% $\leq$ 'F' $\leq$ 60%, and 60% $<$ 'D' $\leq$ 70%, etc. We will return to this issue in §4.

clearly adequate	=	3
neither clearly adequate nor clearly inadequate (i.e., partially adequate)	=	2
clearly inadequate	=	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" (e.g., as in a creative-writing class or an instrumental music class—recall the discussion in §2, above—or perhaps a second-language class) might be graded on a pre-established level of attainment.[6]

3.2 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".) So, the (now, perhaps, misleadingly called) triage[7] theory

[6]For an exception to this 3-point rubric, see §4, esp. n. 15.

[7]I have always thought that 'triage' meant to sort into three categories, but, as my colleague Carl Alphonce has pointed out to me, it doesn't: According to the *Oxford English Dictionary Online* (http://dictionary.oed.com/), '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 excerpt from the 1727 Chambers *Cyclopaedia* entry on "Wool" (p. 377 of the online edition at [http://digicoll.library.wisc.edu/cgi-bin/HistSciTech/HistSciTech-idx?type=turn&entity=HistSciTech000900251034]), 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 15 bales there will be [etc.]. (The '[etc.]' is in the *OED* citation.)

The full passage, though, clearly indicates that the triage is 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.)

Moreover, the *OED*'s next citation, from an 1825 issue of *Gentlemen's Magazine*, 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'.

(Hence my subtitle.)

of grading says that any item to be graded can best be graded on a 4-point scale:[8]

assignment done, and clearly adequate	=	3
assignment done, but only partially adequate	=	2
assignment done, but clearly inadequate	=	1
assignment not done	=	0

0s, 1s, and 3s are intended to be clearly identifiable. Anything not clearly identifiable is a 2. There is some vagueness here: How "clear" must an answer be to be, or not to be, "clearly identifiable"?

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. (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.)[9] 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 §3.3.) 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.[10]

---

[8] Estell (n.d.) has a similar 4-point rubric for engineering education. 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).

[9] This is interpretable as giving the student 1 point for effort. I am not opposed to this. Cf. the discussion in note 10 of a related issue.

[10] 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

However, it is of the greatest importance that you 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" 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 features of grading—namely, dealing with unhappy students—is made much more bearable.

Why 4 grades and not some other number? I argue here that 13 grades ('A', ..., 'D' with + and –, and 'F') is too many distinctions. Walvoord & Anderson (1998) offer several other suggestions: (1) Use 6 grades: 'A', ..., 'F', including 'E', without + or – (these will play a role in my theory, as discussed in §4, below.) (2) Use 4 grades: $\surd$, $\surd+$, $\surd-$, nil (this seems akin to the triage system, but is not singled out for special treatment). (3) Use 3 grades: "outstanding", "competent", "unacceptable" (these particular three levels 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 makes grading "consistent and fair", 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 §5) and convey reasonably precise information to the

correct and incorrect answers (Albert Goldfain, personal communication, 2008). I'm not sure that there's anything wrong with this. Writing 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.

student about what they have learned and what they have not (yet).

3.3 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 is 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 take the exam 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.*

Similar techniques can be applied to other multiple-part assignments. For instance, a programming project that requires a problem definition, a top-down design, documented code, and annotated output might be graded as follows:

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

A philosophy essay that consists of an analysis of an argument can also be graded this way (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 2 premises and a conclusion. Here is one scheme for grading it (refinements should be obvious):

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	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	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 belief	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 required.

Perhaps, however, your essay is less argument-oriented, more open-ended. 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 grades. A simple decision—about whether the student has a generally good grasp of grammar (modulo a comma 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.[11]

[11] Griffin 1998 has a similar 3-part rubric for student essays. (Thanks to Karen M. Wieland for pointing this out.)

3.4 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 2-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*?)

Here are some considerations: 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.[12] Granted, the student who gets the hard part right may not be rewarded as much as they deserve (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 programming project example above, if the instructor feels that documented code is far more important (say, 5 times more) than anything else, the instructor might use this:

[12] As my former math-methods professor, Ann Peskin, advocated.

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

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 problem definition part, *it is also impossible to get more than 10 but less than 15 points on the code part*. You simply explain to the students that their handling of the code was either clearly good, clearly bad, 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", 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.[13]

[13] 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.) On the other hand, use of a fudge factor runs the risk of "students . . . pressing the grader to increase their fudge points" (personal communication, 2008). Thus, it should be used sparingly, if at all.

4 The Triage Theory: Letter Grades

So much for numerical points. How do I convert this to letter grades? 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 only partially adequate
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 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 do many college catalogs) in the sense of "usual", "ordinary", "intermediate".[14] 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). It is, indeed, intermediate between the extremes of "adequate" and "inadequate". (Of course, given the numerical grading scheme (and not counting 0), it is 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".)

What about 'B', you ask? Well, 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

[14] In a "Non-Sequitur" cartoon (28 May 2004), a father says to his daughter, "These grades are way below your potential, Danae. ... How far do you expect to get in this world without knowledge?", to which she replies, "Are you kidding? Ignorance is a great career path." Her father asks, "Uh... how do you figure?"; Danae answers, "Plausible deniability. It's all the rage for corporate boardrooms and heads of state." In the final panel, her father comments to a colleague, "I just realized the world is run by 'C−' students." The colleague replies, "Well, duh... you think things got this way by brainiacs?"

factor	**grade**	**range**	**T=100%**	**width**
$3n$	A	$(17T/18+1)-T$	95–100	$T/18$
$17n/6$	A−	$(8T/9+1)-17T/18$	90–94	$T/18$
$8n/3$	B+	$(5T/6+1)-8T/9$	84–89	$T/18$
$5n/2$	B	$(7T/9+1)-5T/6$	79–83	$T/18$
$7n/3$	B−	$(13T/18+1)-7T/9$	73–78	$T/18$
$13n/6$	C+	$(2T/3+1)-13T/18$	68–72	$T/18$
$2n$	C	$(5T/9+1)-2T/3$	57–67	$T/9$
$5n/3$	C−	$(4T/9+1)-5T/9$	45–56	$T/9$
$4n/3$	D+	$(T/3+1)-4T/9$	34–44	$T/9$
n	D	$(T/6+1)-T/3$	18–33	$T/6$
0	F	$0-T/6$	0–17	$(T/6+1)$

Table 1: From points to letters.

enough, using the mapping described below. (A word of warning: Things get a bit technical at this point.)

Usually, each assignment is worth a multiple of 3 points (e.g., 30 points).[15] To map this into letter grades, let n = the multiple of 3 (in this case, $n = 10$), and let T = the total number of points (so, $T = 3n$; in the example, $T = 30 = 3 * 10$). Note that $3n$ maps to 'A', $2n$ maps to 'C', n maps to 'D', 0 maps to 'F'. Other grades can be interpolated in an evenly-spaced fashion, as shown in Table 1.

This table represents the mapping that I use from the numerical scheme to the letter scheme 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 §5, below.) Incidentally, triage effectively eliminates 'F's except as a message that the student did no work.

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,

[15]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, with no possibility of 2 points. But sometimes a weighting scheme or a very simple problem suggests a point assignment of, say, 0 (i.e., incorrect) or else 1 (i.e., correct).

"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 $5n/2$. (You could, of course, make a different assumption about where 'B' should be interpolated.) With 'B' halfway between 'A' and 'C', I chose to similarly interpolate 'A−' and 'B+' equally spaced between 'A' and 'B'; this results in a raw score of $8n/3$ being mapped to 'B+' and $17n/6$ being mapped to 'A−'.[16] (Again, these letter grades need not be mapped equidistantly; I merely chose to do so.) Similarly, if 'C+' and 'B−' are interpolated equidistantly between 'C' and 'B', they map to raw scores of $13n/6$ and $7n/3$, respectively. Finally, mapping 'D+' and 'C−' equidistantly between 'D' and 'C' maps them to raw scores of $4n/3$ and $5n/3$, respectively. This completes the explanation of the first two columns.

The third column, "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 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 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+' ($= 3.3$)?[17]

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' (i.e., the scores that are $\leq$ a perfect score—clearly an 'A'—but $\geq$ 'A–'). The raw-score endpoints are $17T/18$ and T. Since T is clearly an 'A', the question is whether $17T/18$ should be the lowest 'A' or the highest 'A−'. (This is the same issue discussed in note 5, 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 $17T/18+1$, the lowest 'A−' is $8T/9+1$, etc. This works till we get down to 'D', whose high raw score must (based on my assumptions) be $T/3$. I then assume that the range between 0 and $T/3$ is more-or-less evenly split between 'D' and 'F'; thus, 'D' ranges from $T/6+1$ to $T/3$, and 'F' ranges from 0

[16] The "distance" between the 'A' and 'B' endpoints is $3n - 5n/2 = n/2$, so $1/3$ of the way from 'B' to 'A', which is 'B+', would be $5n/2 + n/6 = 8n/3$. Similarly, $1/3$ of the way from there to 'A', which is 'A−', would be $8n/3 + n/6 = 17n/6$.

[17] 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?

grade	range
A	41–42
A−	38–40
B+	36–37
B	34–35
B−	31–33
C+	29–30
C	24–28
C−	20–23
D+	15–19
D	8–14
F	0–7

Table 2: Letter-grade equivalents when T = 42. Note that I have rounded off some point-values to the nearest whole number; thus, the low end for 'D+', which is precisely $18\frac{2}{3}$, is rounded up to 19, whereas the low end for 'C–' is rounded down to 23 from $23\frac{1}{3}$. Alternatively, one could be strict (rounding all such fractional point-values up) or lenient (rounding them down).

(which has to be its low endpoint) to $T/6$.[18]

I use the "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 42-point programming project above, I would use the letter-grade equivalents shown in Table 2 (T = 42).

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.

[18]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.

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.[19] Alternatively, of course, one can use some non-graded achievement (e.g., attendance or class-participation) to raise a borderline grade.

The fourth column, "$T = 100\%$", shows the mapping of percentages to letter grades 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 the high teens and low 30s). Two things are apparent: This is not a "normal" (or "curved") distribution (see §6), nor is it the "classical" mapping rejected above. As I said in §3.1, I have never understood the classical mapping; it seems completely arbitrary. The Triage Theory at least has a rational basis.

The final column, "width", 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': $T/18$ 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 42-point assignment (here, $T = 42$, $n = T/3 = 14$).

5 On the Subjectivity of Grades

Recall Robert Paul Wolff's "three species of grading", introduced in §1. The second is "evaluation", i.e., "the measuring of a product or performance against an independent *objective* standard of excellence" (Wolff 1969: 59, my emphasis). Clearly, triage is grading as evaluation, but I dispute the objectivity of the standard (in most cases).

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 noted by "criticism", then *calling* it 'good' adds no information. But we teachers are *forced* to summarize our educational

[19]This perhaps violates my desire to curb grade inflation. As I said in footnote 18, so it goes; you may decide otherwise.

factor	**grade**	**range**	***T*=100%**	**width**
42	A	41–42	95–100	2
$39\frac{2}{3}$	A−	38–40	90–94	3
$37\frac{1}{3}$	B+	36–37	84–89	2
35	B	34–35	79–83	2
$32\frac{2}{3}$	B−	31–33	73–78	3
$30\frac{1}{3}$	C+	29–30	68–72	2
28	C	24–28	57–67	5
$23\frac{1}{3}$	C−	20–23	45–56	4
$18\frac{2}{3}$	D+	15–19	34–44	5
14	D	8–14	18–33	7
0	F	0–7	0–17	8

Table 3: From points to letters.

critiques. Hence, triage is inevitable in *our* society, as opposed to the revolutionary one that Wolff wants.

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"; it would also be similar to 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. This can indeed be "ethically troubling".[20] 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

[20]Tanya Christ, personal communication (2008).

is an “informed professional”, it is the *teacher’s* judgment about what the student has learned that is being measured. 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.[21]

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 Grading on a Curve

Although it is not directly relevant to the triage theory, I should 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’, 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.

Grading on a curve is a kind of ranking, 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). Ranking is inevitable once there is “evaluation”. It “performs a function which is

[21]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.

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; emphasis mine.) 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.

7 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.[22] 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 don't understand it at first, but they begin to see how it operates after their first graded assignment. I also provide a grading rubric to accompany all assignments. 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.

8 Summary

The essence of the triage theory of grading is that any item to be graded can and should be graded only as either clearly adequate, clearly inadequate, or neither clearly adequate nor neither 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,

[22] [http://www.cse.buffalo.edu/∼rapaport/howigrade.html].

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?[23]
5. Finally, the grading method is to break complex assignments into simpler parts, each of which is (recursively) graded by triage.

9 Closing Remark

I close with a quote from a "Walnut Cove" cartoon of several years ago:

> Joey (a student): "Mr. Wiggins, can't we curve this F up to a D?"
>
> Mr. Wiggins (his teacher): "Joey, 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."

[23]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).

Acknowledgments

I am grateful to Carl Alphonce, Jonathan Bona, Tanya Christ, Paul V. Gestwicki, Albert Goldfain, Stuart C. Shapiro, Thomas J. Shuell, Paul Vincent Spade, and Karen M. Wieland 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.

References

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

Burke, Christopher (2007), Cartoon, *The Chronicle [of Higher Education] Review* 53(26) (2 March 2007): B13; [http://chronicle.com/weekly/v53/i26/26burke.htm].

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 Time 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 (last accessed, 19 March 2009).

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).

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).

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).

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.