On Assessment

What is education?

To educate is to illuminate the power of ideas.

This, of course, is an ideal – certainly not the last to be articulated, nor perhaps the highest. But putting practicality aside for the moment, how might we take on such a view of education?

Consider the example of planetary orbits. Tying together history, technology, politics, physics and mathematics through a discussion of Kepler's laws is – yes – illuminating. Of course there is no single idea at play here, but consider this: Kepler was perhaps the first astrologer to become an astronomer. He, in contrast to so many of his contemporaries, asked not only *where* a particular planet should be on some future date, but *why* it should be there. In his mind, predicting the positions of celestial bodies for the purpose of casting horoscopes for royal personages was not enough – he wanted to know *why*. That, however, would have to wait until Newton and the application of calculus. And the power of *that* idea!

There is no need to present further historical examples. What *is* important, however, is to move beyond historical ideas and address the issue of the power of a student's *own* ideas. Two semesters of teaching Advanced Problem Solving have shown me that having students write original problems motivates them to learn. They are timid at first – for how could *they* come up with an original problem? But after some success, students are excited to create – and as a result, learn about a particular topic in a more profoundly personal way than they might have otherwise. (Interestingly, the writing of original problems as a teaching tool is common in eighth-grade Japanese mathematics classrooms.[1])

Other assessments are more routine: students are informed that they need to be able to solve linear and quadratic equations, as well as graph linear and quadratic functions. This is perhaps more manageable – the task is well-defined within narrow limits. But it rather different in nature than creating an original piece of work.

Abraham Maslow articulates a similar difference in a discussion of growth theory.[2] Psychologically, why do children grow? "We grow forward when the delights of growth and anxieties of safety are greater than the anxieties of growth and the delights of safety." From an educator's point of view, this implies making the classroom a safe environment for creativity and exploration.

"The opposite of the subjective experience of delight (trusting himself), so far as the child is concerned, is the opinion of other people (love, respect, approval, admiration, reward from

[1]Stigler and Hiebert, *The Teaching Gap,* ISBN 0-684-85274-8, pp. 36–41.
[2]Maslow, Abraham, *Toward a Psychology of Being*, ISBN 0-442-03805-4, p. 47.

others, trusting others other than himself)."[3] Who comes to mind is the student afraid of enrolling in BC Fast-Track because he or she might not earn an A. Or perhaps the student who shuns a difficult course and takes an easier one instead. Grades are just that – opinions of other people.

The assignment of letter grades is not necessary for learning, but is merely practical for other reasons. The assignment of letter grades does nothing to illuminate the power of ideas.

A quick Internet search reveals that the assignment of A–F letter grades is a fairly recent phenomenon, not making its way into high schools until the mid-twentieth century. In the early twentieth century, grades of E (Excellent), S (Satisfactory), N (Needs improvement), and U (Unsatisfactory) were also used.

The question of whether a letter-grade system of evaluation is the best option for a school like IMSA is perhaps a worthwhile one (some argue against grades at all[4]). That is a question which cannot be answered in the short-term, if the question is considered relevant. So the question is this: given that letter grades need to be assigned *this semester,* what approach should be taken?

What do letter grades mean?

The only consensus I have heard, so far, is that earning a grade of A, B, or C means that a student exceeds, meets, or does not meet expectations. Of course, this is reminiscent of E, S, and N. Moreover, a grade of C is passing, so that a student receives credit for a course even if they do not meet expectations.

But this is not what the grades currently mean. Essentially, points are assigned to hundreds of problems given throughout semester, whether on assignments, papers, quizzes, or exams – and an arbitrary weighted sum of these point assignments is converted to a letter. (Of course letters may need to be converted to numbers so that they may be used by an online grade calculator to compute a number which is then converted back to a letter.)

What does this mean about meeting expectations?

Well, nothing really. I would venture to suggest that "meeting expectations" currently means "enough A's, but not too many C's." Perhaps this is politically necessary, but expedient. The performance of our students determines our expectations, rather than the other way around. This is why grades are curved – if grades are too low, then clearly the exam was too difficult.

Now there is a natural give-and-take between evaluating student performance and setting expectations. And, of course, the above remarks are nothing but generalizations. But they illustrate some of the important issues at hand, and may bear fruitful discussion.

[3] Maslow, p. 51.
[4] http://www.alfiekohn.org/teaching/fdtd-g.htm

Moving to more concrete issues, I believe that the assignment of letter grades on exams in BC Fast-Track was, on the whole, successful. Without going into unnecessary detail, the classroom environment was such that the assigned grades were *meaningful* to the students. To give a few examples, I assigned a grade of A+ for truly outstanding work, perhaps only a half-dozen times throughout the entire semester. The students knew this, and so that accolade truly meant something.

Moreover, an A meant something. It was truly rewarding to see the real pride of a student who, used to earning grades in the B range, began to earn the occasional, or perhaps more frequent, A. Admittedly, students who made it to the second semester were essentially guaranteed a grade of no lower than a B−. But this seemed to make an A that much more meaningful.

So student exams had two letter grades on them – one for the skills portion of the exam, and one for the conceptual portion of the exam. No points were assigned, and few comments were made. Students were expected to rework problems on which they made errors.

I bring up this point because I think this system of assigning grades really did motivate students to *learn calculus* rather than *accumulate points.* This is the critical issue: I suggest that the way we assign grades does little to disabuse many students that taking a mathematics course is about accumulating sufficiently many – or losing sufficiently few – points.

Let's take a particular example. The past few semesters, I stopped assigning half-points on assessments. I might forgive a sign error now and then, but too many on a single assessment would warrant a point or two off.

In the past, I simply considered a sign error as a half-point off. And so it was. But consider that without being able to occasionally perform fairly involved calculations, it is not possible to become a successful mathematician. Attention to detail is as important in mathematics as it is in any number of other disciplines, and we try to develop that skill punitively – you *don't* attend to detail, and we *will* take off points.

Of course one might argue that points are given for work well done – but any of us could, I think, agree that when discussing the grading of an exam, it's how many points *off* for a particular type of error that is discussed as much as, or even more than, how many points are *given* for work correctly done.

And so the idea of "partial credit" is born. Perhaps now is not the place to begin *this* discussion, but consider that a student might meet expectations (that is, earn a B) without *ever* having done an entire complex problem on an assessment completely correctly. (Some teachers have even gone so far as to give *no* partial credit.[5]

Why this system of points and partial credit? One may speculate as to its origins, and there is controversy even now about its use on standardized exams.[6] But I cannot help feeling

[5] *On Partial Credit,* Letter to the Editor, MAA Focus, February 2002, p. 17.
[6] http://blogs.villagevoice.com/runninscared/2010/06/test_cheats_on.php

that one function of partial credit is that it allows a teacher to defend the assignment of a particular grade. "*Every* sign error is a half-point off. *That's* why you got a B+ instead of an A−. I have to use the exact same scale for everyone in order to be *fair*."

But doesn't this simply shift the responsibility for the grade onto a rubric? I suggest that many of us would feel competent to take a set of calculus exams – with names removed – and within five or ten minutes, separate out all the A papers. Of course this is subjective – but no less subjective than saying that this problem is worth six points while another is worth ten, or that sign errors are a half-point off, unless, of course, the derivative of the cosine is taken incorrectly, in which case it's a whole point.

Thus the assigning of points is no more "objective" than giving a letter grade. As I'm sure that anyone who has graded a complex word problem based on an assignment of points can attest to. Consider the student who has the entire procedure correct, but because of a few algebra errors, has *no* intermediate calculation correct. The problem is worth ten points. How many points should the student receive?

Well, of course, you say you'd have to see the problem first. But I say, no. The student receives a C. Having *no* intermediate calculations correct demonstrates – regardless of what else – that the student has not met expectations.

So is it possible to avoid points altogether? Perhaps. Consider the following grading system:

Grade	Interpretation
A+	Flawless.
A	Exceeds expectations.
B+	Almost an A.
B	Meets expectations.
C	Does not meet expectations, but passing.
D	Not passing.

Now let's consider this in the context of an exam. The first part of an exam is a skills portion, with, say, ten short problems of roughly equal length. Expectations for this part of the exam are seven problems "essentially" correct, *and* four problems completely correct. These expectations are written on the exam for students to see.

Are these expectations too low? Perhaps. But then an A means eight problems "essentially" correct, with at least five completely correct (please forgive the lapse in logic). Of course we must ask what it means for a problem to be "essentially" correct – but when in doubt, err on behalf of the student. (Students rarely suggest that their scores be lowered.)

Then grading is actually somewhat easier, and grades can be assigned as follows, with the abbreviations EC and CC meaning essentially correct and completely correct, respectively (for simplicity, a grade of C is assigned for all other cases not accounted for):

Grade	Interpretation
A+	All problems completely correct.
A	At least 8 EC, and at least 5 CC.
B+	At least 8 EC and 4 CC, or 7 EC, and at least 5 CC.
B	7 EC and 4 CC.
C	
D	4 or fewer EC.

Now this eliminates the need for partial credit – but does require a judgment as to what "essentially correct" means.

This also makes grading much easier. I would suggest that each problem be marked as "EC", "CC", or left blank. Few comments, if any are necessary. This is the approach I have taken in BC Fast-Track, and it encourages *further* learning as it leaves the student in the position of needing to work through their mistakes.

I would have students keep a section of their notebooks for exams and revisions, and there they can keep their reworked problems, should they choose to do so. Then – as I did in BC Fast-Track – students could visit me periodically with their notebooks and I can take a look at their ongoing progress. This "additional" work, if sufficiently well done, could boost their grade at the end of the semester.

I think this could have the same effect it did in BC Fast-Track – exams were easier and more enjoyable to grade. But there were *more* discussions in my office about reworked exams and sources of error that were initiated by the students themselves, and these discussions were *not* about points, but about *concepts.*

Now what about the part of the exam which is intended to be more conceptual? Let us suppose that there are three problems, roughly comparable in length, and of various difficulties. Then grades might be assigned as follows:

Grade	Interpretation
A+	3 CC
A	3 EC
B	2 EC or 1 CC
C	1 EC
D	0 EC

More details about how this would fit in a classroom environment may be found in a later document. But this system allows for a more qualitative approach to grading. Performance expectations are also clearer, but such expectations depend critically upon the nature of the problems given. Moreover, grades are not assigned punitively, but the emphasis is on doing problems completely *and* correctly.

For an example, below could be a set of ten skills problems and three conceptual questions for a basic assessment on the rules of differentiation. This would a 70-minute assessment. Given expectations for completely correct problems, I think this is reasonable.

Skills questions:

1. Evaluate
$$\lim_{x \to \pi/4} \frac{\tan x - 1}{x - \pi/4}.$$
2. If $f(x) = e^x \cos x$, find $f'(x)$.
3. Find
$$\frac{7}{\sqrt{x^3}}.$$
4. Using the quotient rule, find $\dfrac{d}{dx}\dfrac{x^3}{\sin x}$, simplifying as much as possible.
5. Find the derivative of $f(x) = (\sin \sqrt{x})^2$.
6. Find the equation of the tangent line to $h(x) = \sec(2x)$ at $x = \pi/6$.
7. Using a definition of the derivative, find the derivative of $p(x) = x^2 - x$.
8. Assume that f and g are differentiable functions. Find
$$\frac{d}{dx} f(g(x^2)).$$
9. Find the derivative of $q(x) = x^2 e^x \cot x$.
10. Let f be the greatest integer function. Using the definition of the derivative, determine whether or not the derivative exists at $x = 0$.

Conceptual questions:

1. Using the product rule, find
$$\frac{d}{dx} f(x)(g(x))^{-1}.$$
Explain your result.
2. Suppose that the line $y = 6x + a$ is tangent to both $f(x) = x^2 + b$ and $g(x) = x^3 + 3x$. Find a and b.

3. Suppose that f is a differentiable function. Discuss the following limit:

$$\lim_{h\to 0}\frac{f(x+2h)-f(x-h)}{3h}.$$

There are many further examples of such exams on my BC Fast-Track web sites.

Where does this bring us? Here are some key points as I see them.

1. We should move away from assigning grades punitively.
2. We should reconsider the "point" system of evaluating student performance. Referring to the TIMSS (Third International Mathematics and Science Study): "In our study, teachers were asked what 'main thing' they wanted students to learn from the lesson. Sixty-one percent of U.S. teachers described *skills* they wanted their students to learn. They wanted students to be able to perform a procedure, solve a particular kind of problem, and so on....On the same questionnaire, 73 percent of Japanese teachers said that the main thing they wanted their students to learn from the lesson was to think about things in a new way, such as to see new relationships between mathematical ideas."[7]) A point system reflects the assessment of procedural knowledge.
3. "We can think of all assessment uses as falling into one of two general categories – assessments *FOR* learning and assessments *OF* learning."[8] But why? The distinction is artificial. There are many other ways to compartmentalize assessments, such as timed/untimed, individual/group, skill/conceptual, procedural/relational, short-term/long-term, etc. The main argument for focusing on the "for/or" distinction is its relationship to student motivation – but we are given no context for it. I suggest that our typical IMSA student *is* highly motivated – certainly in relation to the average student in a typical high school classroom.
4. We should consider the assignment of letter grades in general. Right now, it would be impractical to suggest that we have formal written evaluations of each student in each class. But is it desirable? And if so, what resources are necessary to support such a system?
5. We should discuss the assessment of problem-solving.

Will any of these suggestions help to illuminate the power of ideas? I'm not sure. With the current need to assign grades, and their current cultural meaning and importance – especially when it comes to applying to college – there will be the necessary compromises in the classroom. I realize that many suggestions are of the "move away" rather than the "move toward" type. But I suppose that if there is something I am moving toward, it's

[7]Stigler and Hiebert, *The Teaching Gap,* ISBN 0-684-85274-8, pp. 89–90.

[8]Page 29 of Chapter 2 of uncited document distributed to the team.

giving students at all levels more of a BC Fast-Track experience regardless of the depth of content.

This means actively moving toward a classroom environment where earning good grades is subordinate to learning complex concepts. Of course the two are not mutually exclusive – but I'd rather have students earn good grades because they learned, rather than learn in order to get good grades.

Of course many issues brought up in these remarks have been left hanging or only tentatively developed. These brief comments are meant to suggest questions for discussion, not definitive answers.

I can't resist ending with the following challenge from Maslow: "In order to be able to choose in accord with his own nature and develop it, the child must be permitted to retain the subjective experiences of delight and boredom, as *the* criteria of the correct choice for him. The alternative criterion is making the choice in terms of the wish of another person. The Self is lost when this happens."[9] Is it possible to create a mathematics curriculum which can survive this test of course selection?

[9] Maslow, p. 58.