

Criterion Referenced Assessments

W. Blaine Dowler

August 6, 2010

1 Criterion Referenced Assessments Defined

Criterion referenced assessments are those which report student achievement relative to the skills detailed in the curriculum, regardless of peer performance. The most common forms are:

- Letter grades
- Percentage grades
- “Fractional” grades (e.g. “3 out of 4”)
- Skill checklists

The final item on this list, the skill checklist, is probably the least common system. It is, however, the one which makes it easiest for the student to determine which skills need to be improved for the future.

1.1 Letter and Percentage Grades

These are the most common types of grades presented on report cards in public schools. They are most effective in communicating the proportion of the course the student in question has internalized from the course, at least in principle.

So, if both letter grades and percentages are the same basic tool, why are both in use? It’s a question of precision and accuracy. The difference between getting an A and getting a C is fairly obvious and intuitive. The difference between getting 64% and getting 66% is negligible in most cases, but the difference between getting 49% and 51% can make a tremendous difference when

50% is considered the minimum passing grade. Statistically speaking, the uncertainty on a final report card grade may actually exceed 1%, and yet those grades can determine a student's ultimate fate. When statistical error pushes a 51% student into a 49% grade, there's no going back. When one moves to a letter grade system, one can be much more confident that students are getting accurate report card results. There will always be borderline students, but an A/B/C/D/F system has four borders instead of the 99 borders in the percentage system, so students are less likely to be near one.

The down side to this type of grade is that the student doesn't know exactly which skills do and do not need to be refined and improved in the future. One might guess at the relative proportion of skills that need reviewed, but only the most self-aware students will know which particular skills those are.

1.2 Fractional Grades

Fractional grades are most commonly found on individual assignments rather than final report cards. These are the simplest to compute, in that they give the total marks earned relative to the total marks possible. (For example, getting nine questions right on a twelve question multiple choice test usually¹ gives the mark of $\frac{9}{12}$.)

As these marks are typically accompanied by the actual assignment or test which earned the mark, students will often have some idea of what it is they need to work on to improve. However, this information does not come from the mark itself, but from the actual assignment with its individually marked questions. The fractional marks alone are no more or less informative than letter or percentage grades.

1.3 Skill checklists

The final major criterion referenced assessment is the skill checklist. The reason other criterion referenced reporting methods do not inform students of specific skills needing improvement is that the entire curriculum's outcome has been compressed down to a single scale. One needs to maintain multiple scales in order to communicate this information. That's where the checklists come in.

¹In some circumstances, tests are marked "right minus wrong" or with wrong answers being treated somehow differently than blank answers. For example, some tests give four marks for a correct answer, zero for no answer, and negative one mark for an incorrect answer. The idea is to discourage guessing. Instead, one gets a more accurate measure of student confidence levels and risk taking patterns than actual student ability.

A skill checklist reporting scheme is one which lists all outcomes within the curriculum in a checklist, and indicates on an outcome by outcome basis how the student performed. The only method found to be more effective when communicating with students is providing them with detailed, free-form comments on their achievements. Checklists like these are not common in schools, although they can be found in other areas, such as swimming lessons or martial arts instruction.

2 Determining a Final Grade

There are, of course, advantages and disadvantages to this type of grading. The first question is determining how to calculate the grade. We have already discussed the merits and drawbacks of including homework marks in the final grades at all. Beyond that, one must determine the relative weightings of different assessments. How should quizzes compare to exams? What about labs and large reports and projects?

If the primary purpose of giving this grade is to communicate the proportion of the course which a student has internalized, then the seemingly best answers to these questions are counterintuitive and rarely implemented.

The final report card grade should not be a weighted average of different assessments. Rather, it should communicate the curricular outcomes that the student has proven capable of managing by the end of the course. Furthermore, it should be based solely on those curricular outcomes, and not on attendance, behaviour, or other social outcomes which are not explicitly listed in the curriculum.

To compute the grade, one needs to have specific knowledge of the curricular outcomes assessed by each assessment item administered throughout the course. This becomes the source data of a skills checklist. The second counterintuitive point is related to giving credit for each skill: if the student proves competency with a skill at the end of a course, that student should get credit for the skill *regardless* of prior performance. If that student missed the relevant questions on a unit exam but has since learned the material, then *that student has still learned the material, and should receive full credit for doing so*. Once the skills checklist has been completed, an overall grade can be computed from it. However, the student benefits most from the checklist itself. If a teacher wants to best serve all possible recipients of the final grade, then the students should receive the checklist, the overall criterion referenced grade (as a letter, percentage, or whatever system the school has chosen to implement) as well as the relevant norm-referenced information (such as average, standard deviation, and so forth.) This is a tremendous amount of work, and may not be easily

implemented given that most schools mandate report card formats from the administrative level. However, it is still the ideal.

3 Planning: Making the Ideal Feasible

In order to reduce the work required to give the ideal report card down to a manageable level, one must do an immense amount of planning ahead. It may seem counter-intuitive, but one needs to plan the course in reverse.

Begin by writing a final examination which is an accurate representation of the curricular outcomes in the entire course. Keep track of each curricular outcome and the question(s) connected to it. Log them someplace, preferably in some sort of electronic format. Then backtrack the lessons to cover the skills in the appropriate prerequisite order, filling in quizzes and unit exams along the way. This is *not* to say one should “teach to the test.” If the curricular outcome is that “students can use the Pythagorean Theorem to solve for the length of one side of a right angled triangle when given any two other sides,” and the final exam asks students to solve for b in $a^2 + b^2 = c^2$, then the classroom notes and homework should still have an equal share of solving for each of a , b and c . Do *not* simply focus on the one variable that you know is on the test this year, as the students may not be prepared to properly apply the skill if the assignments and applications in subsequent years (including the eventual workplace) require an alternative application of the skill. The electronic assessment item log should include information about all assessment items on all assessments which will contribute to the final report card grade.

The author’s preference for this electronic log when classroom teaching was to create a spreadsheet with a tab for each student, and where each column on a student’s tab represented a single assessment tool. One tab was a master tab, listing all of the assessments from the course to be used for the final report card grades, with both the correlated curricular outcomes and the marks available for each assessment item. Each student’s tab then had a section which was the curricular skills checklist. This was completed automatically through conditional sums on the spreadsheet.² Then, with each assessment administered, a student’s marks were entered on a question by question basis. It took an extra minute to enter the marks this way for each student in the class, but that extra half hour in the day after marking unit exams allowed me the opportunity to hand each and every student a personalized checklist of achievement before the standardized provincial final exam, letting each of them know *exactly* what he or she should focus on while preparing for the final.³ The spreadsheet format

²In both Microsoft Excel and OpenOffice Calc, the SUMIF function serves this purpose quite nicely.

³I did this on my teaching practicums, as I moved into private education before holding

also allows one to adjust the weighting of each assessment item or tool, and to provide students with an immediate update on their current standings, as well as norm referenced classroom data.

4 Upcoming Lessons

Lesson six will conceptually describe mathematical techniques used to analyze individual assessment items in terms of their effectiveness and performance for students in the long term. It will then cover the same material in full mathematical detail for the mathematically inclined. This lesson is typically the highest level material related to assessment that one sees when working on an undergraduate degree in education. Lessons seven, eight and nine continue from there, detailing some advanced models and techniques, again covering concepts first in detail, and then following up with the full mathematical glory in sections which can be omitted without making later lessons difficult to follow.

a regular classroom position for a full year. My cooperating/supervising teacher, who was formally in charge of the class, later informed me that the classes I worked with performed 10% higher than the provincial averages on those standardized assessments. Quality checklists shared with students before high stakes exams can turn into very effective study guides.