Introduction

The goal of this document is to provide a concise summary of the underlying principles of assessment and ideas for alternative methods. May this paper serve as a framework to help instructors design or adopt alternative assessment strategies even if class size increases.

What is assessment?

Assessment is a term applied to a wide range of approaches used to measure what our students have learned. Contemporary theories of learning and knowing emphasize the way knowledge is represented, organized, and processed in the mind (Knowing What Students Know, 2001) The authors of this study go on to explain that estimates of how people organize information in long-term memory are likely to be more important than estimates of working (short term) memory capacity: "understanding the contents of long-term memory is especially critical for determining what people know, how they know it, and how they are able to use that knowledge to answer questions, solve problems, and engaged in additional learning.” Much of what one knows is domain- and task-specific and organized into smaller chunks known as schemas. Assessment should evaluate what schemas an individual has and under what circumstances he/she regards the information as relevant. This evaluation should include how a person organizes acquired information, encompassing both strategies for problem solving and ways of chunking relevant information into manageable units.

Experts (instructors) and novices (our students)

The importance of evaluating knowledge structures comes from research on expertise (Knowing what students know, 2001). Experts in a subject domain typically organize factual and procedural knowledge into schemas that support pattern recognition and the rapid retrieval and application of knowledge. Novices, on the other hand show varying abilities to organize information and to see patterns. Our goal as experts is to help novices acquire the skills to organize factual and procedural knowledge in our subject area. As instructors we intuitively or explicitly do that, but in larger classes our methods need to be even more explicit. We also need to consciously articulate to our students as many strategies as possible to reach people with diverse learning styles who hear and process information differently than we do.
What is assessment really trying to do?

Don't confuse feedback and evaluation of our students' understanding

Formative assessment provides feedback to both instructor and students on how well students understand course material. The outcome from formative assessment should be to improve individual student learning as well as improve/alter teaching strategies to focus on weak areas. Formative assessment is not part of the final mark, but helps both you and the students know where they are and what instructors and students need to focus on.

Summative assessment provides a grade/mark reflecting student achievement. Error in summative assessment is now perceived by students as punishment not informative, Biggs (1999:143) illustrates the difference between formative and summative assessment with this analogy "When the chef tastes the sauce it is formative; when the customer tastes, it is summative".

This section on assessment focuses on summative assessment; however, formative assessment is also important. Field Tested Learning Assessment Guide (FLAG) has some good ideas [http://flaguide.org/].

What is the basis for distinguishing between students?

Before we choose assessment strategies, we also need to consider what is the basis for the outcome of our evaluation? Do we design projects, tests, assignments to evaluate students’ ability to deal with certain types of knowledge and/or skills (criteria based assessment) or is our assessment based on giving tests, projects, etc. and then ensuring that the resulting marks fit a normal distribution with a pre-set percent As Bs, Cs etc (norm referenced assessment)? Both types of assessment can provide the same distribution of marks, but there are significant and fundamental differences in the construction of the assessment-and the impact on student motivation as well as their perceived sense of "fairness" of the evaluation.

Criterion referenced assessment (CRA) is the standards model designed to assess changes in performance as a result of learning, for the purpose of seeing what and how well something has been learned. The result is to see how well students have learned course objectives. I would argue that at university our assessment should be CRA based on predetermined standards from clear course goals and objectives rationalized for the course and program. Note CRA should not be confused with mastery learning-just because the objectives are explicitly stated does not mean all students can achieve them to the same extent-not all students will get an A!

Norm referenced assessment (NRA) is the measurement model designed to assess stable characteristics of individuals for the purpose of comparing them with each other. The result is to rank order students. NRA does not indicate how well students have learned something, just how they compare to other students.
Comparison of Norm (NRA) and Criterion (CRA) Referenced Assessment

<table>
<thead>
<tr>
<th></th>
<th>NRA</th>
<th>CRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standards set</td>
<td>After teaching &amp; test given</td>
<td>Before teaching takes place</td>
</tr>
<tr>
<td>Results expressed as</td>
<td>Comparison between students</td>
<td>How well student performance matches set criteria</td>
</tr>
<tr>
<td>Results</td>
<td>Dependent on other students</td>
<td>Independent of other students</td>
</tr>
<tr>
<td>Judgment based on</td>
<td>People (rank order)</td>
<td>Performance (standards)</td>
</tr>
</tbody>
</table>

What is the difference between normative and criterion based assessment?

» Try this quiz for clarification of the difference between CRA and NRA (Biggs, 1999:149).

1. I am assessing two students in a CRA system and note that I have assigned Robert a "B" and Susan an "A". I re-look at Robert's paper and decide his is as good as Susan's, so I give them both an A. Is this now NRA (comparing people) or CRA (judging on standards)? [answer]

2. I am intending to structure my evaluation using a CRA system and in deciding what standards are reasonable to set for the grading categories, I look at last year's distributions of As, Bs, Cs, Ds, and Fs and adjust the difficulty of the tasks so I am likely to obtain a similar result this year. NRA or CRA? Why [answer]

How does class size impact assessment?

Corollary: What is large?

If you usually teach small seminar courses of about 20 students, a class of 60 or 100 will seem large and force you to reconsider how to evaluate your students. Scaling up certain types of assessment (e.g., oral presentations, large research reports) even with the money for marking assistance is not always realistic considering constraints in time, space, and the availability of sufficient qualified markers. In addition, multiple markers introduce problems including the need for you to train the markers and the real or perceived potential inconsistency between markers.

Marking consistency may be improved significantly with assessment rubrics and marking workshops for markers. Even when you achieve marking consistency, you need to consider various strategies to provide students with useful feedback. Using rubrics save markers time writing out comments, but students need to be taught how to read the rubrics-do not assume words such as synthesis, logical, or appropriate sources of evidence are understood.
A class of about 100-300 is large; however mega classes of over 1000 pose their own set of issues and should be considered qualitatively as well as quantitatively different. The potential 'distress level' of disgruntled students before, during, and after tests (or major projects or even small assignments) is typically about 10% of the students: that is 10 students in a "large" class of 100, but 100 students in a mega class. Have a clearly defined procedure for help sessions, remedial support, alternative makeup tests, review/remark requests, etc. [This page will compile procedures and post them to use as a guide for your own course.]

**Align assessment with course goals**

Effective assessment begins with clear goals. Before we can assess how well our students are learning we need to identify and clarify what we are trying to teach. At first glance the question of what we are teaching seems obvious with answers such as "I am teaching introductory molecular biology", or "the nineteenth century British novel" or "advanced organic chemistry". However, we tend to define our instructional goals in terms of course content, rather than articulate the underlying skills and competencies we hope our students learn through the course content.

**Underlying goals of evaluation for a particular class**

Before launching into a series of ideas on how to evaluate step back and ask yourself some questions about the underlying goals in a particular course. Your answers will impact the type of evaluation you will want to use with your students and the options available in a large class setting.

**What content and skills are this course designed to achieve as part of the students' program?** (Evaluation needs to reinforce both the content and skills.)

- What do you want your students to have learned by the end of the term in terms of content, higher order thinking, writing, numeracy, and problem solving skills and discipline-specific techniques?

- What do instructors in advanced courses assume students have learned in this course in terms of content, higher order thinking skills, and discipline-specific skills and techniques?

Introductory courses tend to emphasize objective factual information as a foundation for the discipline. Evaluation methods, such as multiple choice tests, are well suited to assess objective knowledge and are efficient in large classes. More advanced courses expect students to apply knowledge and synthesize ideas which require more open-ended types of evaluation. Short projects or larger group projects in large advanced classes make grading more manageable than large individual research projects yet still retain the goal of assessing synthetic thought. Vary the format of a final project from a large research paper to individual or group poster sessions, web pages, or perhaps a lecture/seminar on a given topic so that the marking and feedback from marking guideline rubrics may be more efficacious in large and small classes.
Where does your course fit into the students' programmes?
This question reflects issues related to course level from introductory to advanced, students' intellectual maturity, and students' inherent interest in the subject area, all of which will affect student motivation in the course and students' understanding of the perceived relevance of the material.

- Is the course in the students' area of specialization? Whether students are taking a required or elected course outside their area of specialization means students are less familiar with the subject-specific factual and procedural knowledge. As an instructor you will need to make explicit the strategies you use that support pattern recognition to help these students learn their own means of rapid retrieval and application of knowledge.

- Are most of your students in first, second, or upper year? An introductory class with a significant percentage of upper year students is tougher to teach as you are faced with a more pronounced mix of intellectual maturity. Evaluations stressing primarily objective recall rather than problem solving or critical thinking favour the first year students and may not truly reflect the advanced student's knowledge. As we push our students to higher levels of understanding and intellectual development, we need to be careful neither to set learning goals so high that they are beyond the reach of most students nor too low so we don't challenge the students ready to make the intellectual jump. We can stretch more of our students by setting incremental goals.

- Is your course a program prerequisite or an elective? Students who elect to take a course tend to be more interested, while students who "have to take" the course because it is a program requirement, may not be inherently interested (although they want to do well). Students do better when they understand and share the goals for learning. In large classes we need to work harder and be more creative to engage students with diverse backgrounds and goals. In larger classes, the subpopulation of unengaged students may be a sizable group: for example, while 10% of a 30 student class is only 3 students, 10% of 100- or 1000- or 2000-student class is 10, 100, 200 students respectively.

How does your teaching style and evaluation match what you expect the students to learn?

Corollary: What do your students think is important and is it the same as what you think is important?

Students learn what they think is important based on how they will be evaluated- Ramsden (1992:187) coined the term "backwash" referring to the implicit message we give students about what is really important. If we give objective tests, that is what is important, not insightful synthetic thought. The type of evaluation needs to match our real goals!
As instructors, we see the curriculum objectives as central to an aligned course/programme. However, from our students' point of view, assessment always defines the actual curriculum. If the bulk of the course marks is based on remembering factual information, that is what students see as important, not critical thinking.

Strive to use a variety of methods within a single course to assess different aspects of student learning. Sounds great, but how to achieve that variety in large and mega classes?

Resources

1 - Field-tested Learning Assessment Guide (FLAG) for science, math, engineering, and technology instructors (STEM) (http://flaguide.org/)

FLAG offers broadly applicable, self-contained modular classroom assessment techniques (CATs) and discipline-specific tools for instructors interested in new approaches to evaluate student learning, attitudes and performance. FLAG also contains an assessment primer for a concise explanation of what we know about assessment, a section to help select the most appropriate assessment technique for your course goals, and other resources.

Assessment Primer from the Field-Tested Learning Assessment Guide, is an excellent discussion about the various issues underlying assessment covered superficially in this summary.

2 - McAlpine, M., CAA Centre (http://www.caacentre.ac.uk/) [accessed Nov. 2003]

This site was designed to provide information and guidance on the use of computer-assisted assessment (CAA) in higher education. The folder "Resources" has a series of Bluepapers as well as useful guidance on the construction of tests. The folder "Blueprint" refers to the text published in November, 2003, published by Routledge Falmer, ISBN 0415287030.

Bluepapers series was developed to focus upon a CAA related issue in more depth than previously allowed. Each paper is available for free download but should not be stored or distributed electronically as outlined within the copyright statements.

- **Bluepaper 1:** Principles of Assessment by Mhairi McAlpine This paper reviews the context and tensions within educational assessment and refreshes memory of the purposes of assessment.

- **Bluepaper 2:** A Summary of Methods of Item Analysis by Mhairi McAlpine Item analysis is a method of gauging the quality of an examination by looking at its constituent parts (items). The primary purpose of most examinations in higher education is that of a measurement tool, for assessing the achievements of the examination candidates, and thus how future learning can be supported and directed.
This paper details the methods of item analysis for this purpose, and also considers how they might be used for the wider functions given above. Download Bluepaper 2 as PDF File (11514K) at http://www.caacentre.ac.uk/dldocs/BP2final.pdf

- **Bluepaper 3**: Design Requirements of a Databank by Mhairi McAlpine This Bluepaper gives an insight into some of the concerns and considerations that must be taken into account when designing an item banking system. Download Bluepaper 3 as PDF File (10526K) at http://www.caacentre.ac.uk/dldocs/BP3final.pdf

---

**3 - GLOW** ([http://www.glow.ac.uk/](http://www.glow.ac.uk/))

Web-based Assessment: A new conceptual framework for supporting student progression and achievement. This site has some very useful information if you want to use online assessment: getting started and evaluation of web-based authoring tools.

---

**4 - Anderson, R.S. and Puckett, J.B. (fall) 2003** *Assessing Student's Problem-Solving Assignments*, New Directions for Teaching and Learning 95 pp 81-87. (Available online through U of T library e-journals)

---

**5 - Teaching Goals Inventory** (Angelo and Cross, 1993:20-22) Online survey for instructors to analyze their teaching goals http://www.uiowa.edu/~centeach/tgi/background.html

The Teaching Goals Inventory (TGI) is a self-assessment of instructional goals. Its purpose is (1) to help university instructors become more aware of what they want to accomplish in individual courses; (2) to help faculty locate Classroom Assessment Techniques they can adapt and use to assess how well they are achieving their teaching and learning goals.

---

**References**


1. In question one, Susan's and Robert's performance are compared to check the consistency of the assessment, not to rank them for grades. In CRA a student's performance is compared to predetermined standards to see which grading category applies. In this example the instructor's initial judgment of Robert's performance was not consistent with the rubric/standards. The standards themselves were not altered. This example of assessment is CRA.

2. In the second example the standards for assessment were defined before grading began. The fact that the standards were defined in part from norm-referenced data is not relevant. At the end of the unit each student's performance would be compared to those preset standards not with each other: hence CRA.