Assessment in Higher Education

N.J. Rao

1. Introduction

Assessment is defined as a measure of performance. Evaluation is an interpretation of assessment. Educational assessment is the process of documenting, usually in measurable terms, knowledge, skills, attitudes and beliefs. Assessment is a mechanism for providing instructors with data for improving their teaching methods and for guiding and motivating students to be actively involved in their own learning. As such, assessment provides important feedback to both instructors and students. Assessment gives us essential information about what our students are learning and about the extent to which we are meeting our teaching goals. But the true power of assessment comes in using it to give feedback *to students*. Improving the quality of learning in a course involves not just determining to what extent students have mastered course content *at the end of the course*; improving the quality of learning also involves determining to what extent students are mastering content *throughout the course*. Thus in addition to providing the instructors with valuable information about our students' learning, assessment should assist the students in diagnosing *their own* learning. That is, assessment should help students "become more effective, self-assessing, self-directed learners." (Angelo & Cross, 1993)

There is considerable evidence showing that assessment drives student learning. More than anything else, our assessment tools tell students what we consider to be important. They will learn what we guide them to learn through our assessments. Traditional testing methods have been limited measures of student learning, and equally important, of limited value for guiding student learning. These methods are often inconsistent with the increasing emphasis being placed on the ability of students to think analytically, to understand and communicate at both detailed and "big picture" levels, and to acquire life-long skills that permit continuous adaptation to workplaces that are in constant flux. Moreover, because assessment is in many respects the glue that links the components of a course - its content, instructional methods, and skills development - changes in the structure of a course require coordinated changes in assessment.

2. Types of Assessment

Assessment is often divided for the sake of convenience using the following distinctions:

1. Formative and summative
2. Objective and subjective
3. Referencing (criterion-referenced, norm-referenced, and ipsative)
4. Informal and formal

Formative and summative: Assessment is often divided into formative and summative categories for the purpose of considering different objectives for assessment practices.

Summative assessment is intended to measure learning outcomes and report those outcomes to students, parents, and administrators. In an educational setting, it generally occurs at the conclusion of a class, course, semester, or academic year. In the context of a course, summative assessments are typically used to assign students a course grade. It is also referred to in a learning context as "assessment of learning". *Performance-based assessment* is similar to

summative assessment, as it focuses on achievement. A well-defined task is identified and students are asked to create, produce, or do something, often in settings that involve real-world application of knowledge and skills. Proficiency is demonstrated by providing an extended response. Performance formats are further differentiated into products and performances. The performance may result in a product, such as a painting, portfolio, paper, or exhibition, or it may consist of a performance, such as a speech, athletic skill, musical recital, or reading.

Formative assessment is generally carried out throughout a course or project. In an educational setting, formative assessment is used by teachers to consider approaches to teaching and next steps for individual learners and the class, and would not necessarily be used for grading purposes. Formative assessment, also referred to as "educative assessment" or "assessment for learning," is used to aid learning. Assessment for learning is defined as "all those activities undertaken by teachers and/or students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged" (Black and William 2004). The key features of assessment for learning are

High quality interactions: Classroom assessment involves high quality interactions, based on thoughtful questions, careful listening and reflective responses;

Involving students in their learning: Students and instructors are fully involved in deciding next steps in their learning and identifying who can help.

Feedback: Students and instructors are given timely feedback about the quality of their work and how to make it better.

Sharing criteria: Students and instructors are clear about what is learned and what success would be like.

A common form of formative assessment is "*diagnostic assessment".* Diagnostic assessment measures a student's current knowledge and skills for the purpose of identifying a suitable program of learning. "*Self-assessment"* is a form of diagnostic assessment which involves students assessing themselves. "*Forward-looking assessment"* asks those being assessed to consider themselves in hypothetical future situations.

Objective and subjective: Assessment (either summative or formative) is often categorized as either objective (selection) or subjective (supply). Objective or selection type assessment is a form of questioning which has a single correct answer. Subjective assessment is a form of questioning which may have more than one correct answer (or more than one way of expressing the correct answer). There are various types of objective and subjective questions. Objective question types include true/false answers, multiple choice, multiple-response and matching questions. Subjective questions include extended-response questions, essays, hypothesizing, creating plans etc. Objective assessment is well suited to the increasingly popular computerized or online assessment format.

Some have argued that the distinction between objective and subjective assessments is neither useful nor accurate because, in reality, there is no such thing as "objective" assessment. In fact,

all assessments are created with inherent biases built into decisions about relevant subject matter and content, as well as cultural (class, ethnic, and gender) biases.

Referencing: Test results can be compared against an established criterion, or against the performance of other students, or against previous performance. *Criterion-referenced assessment,* typically using a criterion-referenced test, as the name implies, occurs when candidates are measured against defined (and objective) criteria. Criterion-referenced assessment is often, but not always, used to establish a person's competence (whether s/he can do something). The best known example of criterion-referenced assessment is the driving test, when learner drivers are measured against a range of explicit criteria (such as "Not endangering other road users").

Norm-referenced assessment (colloquially known as "grading on the curve"), typically using a norm-referenced test, is not measured against defined criteria. This type of assessment is relative to the student body undertaking the assessment. It is most appropriate when one wishes to make comparisons across large numbers of students or important decisions regarding student placement and advancement. Norm-referenced measures are designed to compare students (i.e., disperse average student scores along a bell curve, with some students performing very well, most performing average, and a few performing poorly). The IQ test is the best known example of norm-referenced assessment. Many entrance tests (to prestigious schools or universities) such as CET, GATE, CAT, SAT and GRE are norm-referenced, permitting a fixed proportion of students to pass ("passing" in this context means being accepted into the school or university rather than an explicit level of ability). This means that standards may vary from year to year, depending on the quality of the cohort; criterion-referenced assessment does not vary from year to year (unless the criteria change).

Ipsative assessment is self comparison either in the same domain over time, or comparative to other domains within the same student.

Informal and Formal: Assessment can be formal or informal. Formal assessment usually implicates a written document, such as a test, quiz, or paper. A formal assessment is given a numerical score or grade based on student performance, whereas an informal assessment does not contribute to a student's final grade. An informal assessment usually occurs in a more casual manner and may include observation, inventories, checklists, rating scales, rubrics, performance and portfolio assessments, participation, peer and self evaluation, and discussion.

Internal and External: Internal assessment is set and marked by the institution (i.e. teachers). Students get the mark and feedback regarding the assessment. External assessment is set by the governing body, and is marked by non-biased personnel. With external assessment, students only receive a mark. Therefore, they have no idea how they actually performed (i.e. what questions they wrote correctly)

3. Standards and Quality

In general, high-quality assessments are considered those with a high level of reliability and validity.

Reliability relates to the consistency of an assessment. A reliable assessment is one which consistently achieves the same results with the similar cohorts of students. Factors that affect

reliability include systematic errors like insufficient competency, cognitive level coverage, ambiguous questions, too many options within a question paper, vague marking instructions and poorly trained markers, random errors like presence of learner specific factors (motivation, concentration, fatigue, boredom, etc.), and assessment environment specific factors (noise distractions, poor lighting, uncomfortable room temperature). The two forms of reliability that typically are considered in classroom assessment are:

1. Interrater Reliability: It refers to the consistency of scores that are assigned to two independent raters.
2. Intrarater Reliability: It refers to the consistency of scores that are assigned by the same rater at different points in time.

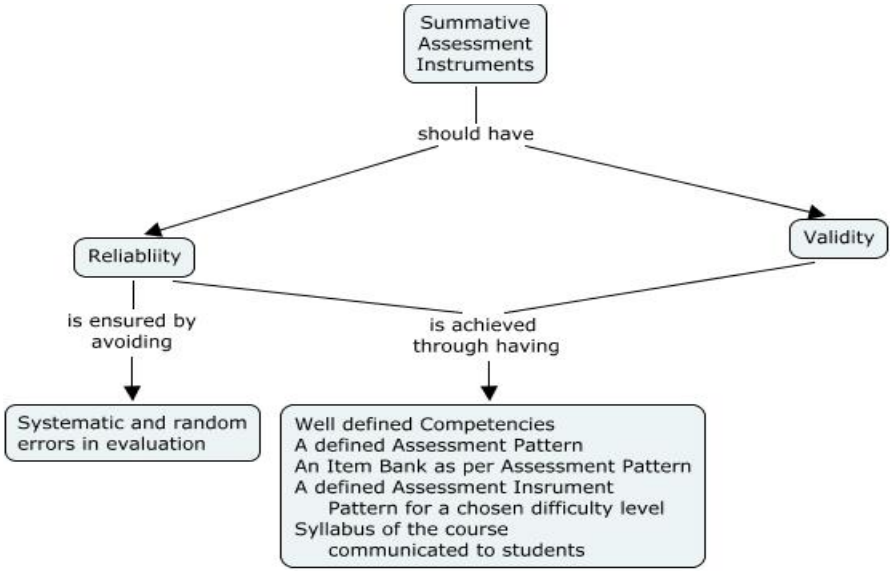
Validity: A valid assessment is one which measures what it is intended to measure. For example, it would not be valid to assess ability to design a circuit through asking the student to explain the theory of devices used in the circuit. The measures of validity relevant to courses in formal engineering programs are

1. Curriculum Validity: The outcome must be defined and worth achieving.
2. Content Validity: The assessment must reflect the content and balance of the teaching and learning.
3. Face Validity: The assessment must seem credible to students and other stakeholders.

A good assessment has both validity and reliability, plus the other quality attributes noted above for a specific context and purpose.

Summative assessment instrument can be both valid and reliable by having

- Well defined Competencies
- A defined Assessment Pattern
- An Item Bank as per Assessment Pattern
- A defined Assessment Instrument
- Pattern for a chosen difficulty level
- Syllabus of the course communicated to students



5. Written Test Items

Written test items are given in quizzes, class tests, assignments and final examinations or in the forms of reports students are required to generate. The test items can be **supply-type items** or the **selection-type items**. For each specific learning outcome a number of test items can be prepared for assessment before instruction, during instruction and assessment at the end of instruction.

Selection-type items are highly objective. The questions have unique answers. Tests that include only selection-type items make it possible to directly compare student accomplishment. For example, it is possible to compare the performance of students within one class to students in a different class, or students under one instructor with those under another instructor. By using selection-type items, the instructor can test on many more areas of knowledge in a given time than could be done by requiring the student to supply written responses. Typically Selection Type items are used to assess lower cognitive levels, as questions asked in lower cognitive levels usually have a unique answer (consider definitions, lists, formulae, etc). Evaluation of students' performance in selection type test items is easier to automate. Examples of Selection Type questions are:

- True/False
- Multiple Choice (with single or multiple answers)
- Matching Blocks
- Rearrangement
- Checklists
- Rating Scales

Supply-type item may be required where a selection-type cannot be devised to properly measure student knowledge, typically higher cognitive levels. The supply-type requires the students to organize their knowledge. It demands an ability to express ideas that is not required for a selection-type item.

This type item is valuable in measuring the students' generalized understanding of a subject. On the other hand, a supply-type item may evaluate the students' ability to innovate (cognitive level: Create) rather than their specific knowledge of the subject matter. The main disadvantage of supply-type tests is that they cannot be graded with uniformity, which presents great difficulties in automating the evaluation process. The same test graded by different instructors would probably be assigned different scores. Even the same test graded by the same instructor on consecutive days might be assigned altogether different scores. Therefore, standardization of evaluation is not always possible. Still another disadvantage of a supply-type test is the time required by the student to complete it and the time required by the instructor to grade it. Supply-type test items are necessary when there is a need to thoroughly determine the ability of the student to use higher cognitive levels in the course. Answers to supply-type questions may not be unique. Examples of Supply Type questions are:

- Completion
- Sketch

- Labeling
- Short Answers
- Structured Response
- Viva Voce
- Numerical Questions
- Detailed Answers
- Critical Analysis

Sample solutions to Test Items: Solutions to the representative written test items should be prepared to reflect what the instructor considers as an appropriate answer and how the selected program outcomes are addressed. Solutions help students to follow good practices in writing their solutions to test items.

Performance Test Items: Performance test items include experiments to be conducted, simulations to be done, projects, prototypes to be built, things/objects/equipment/physical samples to be tested, discussions, presentations, field studies etc. They will involve working with devices, performing, discussions, and report generation. These activity items can be very definitive or open ended. The evaluation of performance of students in activity based test items is likely to be qualitative. For every type of test item it becomes necessary to define rubrics.

5. Rubrics

A **rubric** is a scoring tool for subjective assessments. It is a set of criteria and standards linked to learning objectives that is used to assess a student's performance on papers, projects, essays, and other assignments. Rubrics allow for standardized evaluation according to specified criteria, making grading simpler and more transparent. (Wikipedia)

The rubric is an attempt to delineate consistent assessment criteria for grading. Because the criteria are public, a rubric allows teachers and students alike to evaluate criteria, which can be complex and subjective. Rubrics also provide a basis for self-evaluation, reflection, and peer review. It is aimed at accurate and fair assessment, fostering understanding and indicating the way to proceed with subsequent learning/teaching.

Instructors who use rubrics to evaluate student performance should share the rubrics with students at the time the assignment is made. Sharing rubrics enables students to do the assigned tasks as expected and provides alignment between competencies and assessment. Rubrics help students become thoughtful evaluators of their own and others' work and reduce the amount of time teachers spend evaluating student work (Goodrich, 1996).

Components of a Rubric: The components of a rubric are Attributes, Scoring, and Descriptors. Each rubric will have a set of attributes which constitute the criteria for evaluation. These can be identified by the instructor or jointly by the instructor and students. www.rubistar.4teachers.org presents rubrics for a very large number of subjective assessments. Examples of attributes of a rubric for presentation are

- Content
- Enthusiasm
- Suggested Elements

- Required Elements
- Overall Impression

Two samples of rubrics indicating thee attributes, scoring and descriptors from www.rubistar.4teachers.org are given in Annexure 1

6. Summary

Learning by students in higher education is dictated the actual assessments (for example final examination papers of last five years) or assessments as perceived by students (perception of what the teachers considers important and related knowledge about the teacher transferred from senior students). Unfortunately many teachers at higher educational institutions still complain about the preoccupation of students with examinations, and in their view "learning well" is at variance with assessment. In fact the single most important intervention to improve the quality of learning is assessment. The quality of learning can be measured by the quality of assessment.

While formative assessment can be very important in providing effective feedback to both teachers and students, in practice it is not always effective. Students everywhere do not exert themselves if an assessment instrument is not an element of summative assessment. The only method that would work is to have summative assessment across several assessments spread through the course.

Too many assessments may lead to taking fun out of learning. Besides, any increase in the number of assessments leads to increase in the workload of teacher, more so if the number of students is large.

When higher orders of learning (analysis, evaluate and create cognitive levels of Bloom) are being assessed, the responses of students will not be unique nor there are unique correct answers. When the student strength in a course is large, it can lead to significant increase in the time for evaluation by the teacher. Therefore, assessment is a major exercise in balancing the quality of learning, work load of the teacher and the fun in learning.

References:

1. Angelo, T. A. & Cross, K. P., Classroom assessment techniques: A handbook for college teachers, San Francisco, Jossey-Bass, 1993.
2. Nlack, P., William, D., The Formative Purpose: Assessment must first promote Learning in Towards Coherence Between Classroom Assessment and Accountability – 103 Yearbook of the National Society for the Study of Education (Ed: M. Wilson), pp. 20-50, Chicago: University of Chicago Press for the NSEE
3. Linn, R. L., *Measurement and assessment in teaching* (7th ed.). Englewood Cliffs, NJ: Merrill, 1995.
4. Goodrich, H. (1996): "Understanding Rubrics." Educational Leadership, 54 (4), 14-18.
5. Rubrics: http://rubistar.4teachers.org/

Annexure 1: Sample rubrics from www.rubistart.4techers.org

Lab Report : Chemical Engineering Laboratory

CATEGORY	4	3	2	1
Lab Preparation and Performance	Data sheets are prepared. Student demonstrates superior knowledge of the experimental method and theory. Student properly executes experiment, and properly cleans up.	Data sheets are prepared. Student demonstrates above average knowledge of the experimental method and theory. Student properly executes experiment, and properly cleans up.	Data sheets are not prepared. Student demonstrates adequate knowledge of the experimental method and theory. Student has problems executing the experiment, and does not clean up adequately.	Data sheets are not prepared. Student demonstrates poor knowledge of the experimental method and theory. Student has a large number of problems executing the experiment, and does not clean up.
Background and Theory	Student gives a superior overview of the experimental process, and underlying fundamental principles. Several examples of industrial applications are provided. Text is reference appropriately.	Student gives an above average overview of the experimental process, and underlying fundamental principles. Few examples of industrial applications are provided. Text is reference appropriately.	Student gives an adequate overview of the experimental process, and underlying fundamental principles. One or no examples of industrial applications are provided. A few statements require a reference.	Student gives a poor overview of the experimental process, and underlying fundamental principles. One or no examples of industrial applications are provided. A large number of statements require a reference.
Sample Calculations	Sample calculations clearly demonstrate how data is transformed into actual results. Each step is explained, and any outside constants are cited. All spreadsheets and graphs are included.	Sample calculations somewhat demonstrate how data is transformed into actual results. All steps are not explained, though any outside constants are cited. All spreadsheets and graphs are included.	Sample calculations somewhat demonstrate how data is transformed into actual results. All steps are not explained, and any outside constants are not cited. Some spreadsheets and graphs are included.	Sample calculations do not demonstrate how data is transformed into actual results. All steps are not explained, and any outside constants are not cited. No spreadsheets and graphs are included.
Results and Discussion	Pertinent figures and tables are included in the body. Student gives a superior discussion of the results presented in the figures and tables. A comparision of acquired results with literature data is given. Abstract clearly summarizes the experiment, names the important process variables, and is an appropriate length. The conclusion clearly summarizes all key quantitative and qualitative results of the study, and is an	Pertinent figures and tables are included in the body. Student gives an above average discussion of the results presented in the figures and tables. A comparision of acquired results with literature data may or may not be given. Abstract somewhat summarizes the experiment, names the important process variables, and is an appropriate length. The conclusion somewhat summarizes all key quantitative and	Pertinent figures and tables are included in the body. Student gives an adequate discussion of the results presented in the figures and tables. A comparision of acquired results with literature data may or may not be given. Abstract somewhat summarizes the experiment, names the important process variables, and is not an appropriate length. The conclusion somewhat summarizes all key quantitative and	Pertinent figures and tables are included in the body. Student gives a poor discussion of the results presented in the figures and tables. A comparision of acquired results with literature data is not given. Abstract does not summarize the experiment, name process variables, and is not an appropriate length. The conclusion does not summarize all key quantitative and qualitative results of the study, and is not an appropriate

	appropriate length.	qualitative results of the study, and is an appropriate length.	qualitative results of the study, and is not an appropriate length.	length.
Statistical Error Analysis	Derivation of error equations are explained during each step. Opening paragraphs gives a superior discussion of why specific equations were selected. Concluding paragraphs give a superior description of absolute and percentage error in the results.	Derivation of error equations are explained during each step. Opening paragraphs gives an above average discussion of why specific equations were selected. Concluding paragraphs give an above average description of absolute and percentage error in the results.	Derivation of error equations are not explained well or completely. Opening paragraphs gives an adequate discussion of why specific equations were selected. Concluding paragraphs give an adequate description of absolute and percentage error in the results.	Derivation of error equations are not explained at all. Opening paragraphs gives a poor discussion of why specific equations were selected. Concluding paragraphs give a poor description of absolute and percentage error in the results.
Organization and Clarity	Cover includes the experiment number, title, name of group leader, date of experiment, and date of submission. All sections are clearly labelled. All formatting is correct, and the report has a neat appearance overall. All figures and tables are in the correct format. The student has a superior grasp of standard written english with no spelling or grammatical errors.	Cover includes the experiment number, title, name of group leader, date of experiment, and date of submission. Sections are adequately labelled. There are some formatting errors, though the report has a neat appearance overall. All figures and tables are in the correct format. The student has an above average grasp of standard written english with minimal spelling or grammatical errors.	Cover does not include the experiment number, title, name of group leader, date of experiment, and date of submission. Sections are hard to distinguish despite being adequately labelled. There are several formatting errors, and the report has an adequate appearance overall. Some figures and tables are in the correct format. The student has an adequate grasp of standard written english with several spelling or grammatical errors.	Cover does not include the experiment number, title, name of group leader, date of experiment, and date of submission. Sections are hard to distinguish and not labelled. There are several formatting errors, and the report has a poor appearance overall. Few figures and tables are in the correct format. The student has a poor grasp of standard written english with a large number of spelling or grammatical errors.

Building A Structure : Design Engineering Project

CATEGORY	4	3	2	1
Information Gathering	Accurate information taken from several sources (interview people, internet, retail outlets). Ideas and sources documented.	Accurate information taken from some sources.2 of 3 (interview people, internet, retail outlets). Ideas and sources documented.	Accurate information taken from one other source than self. Ideas and sources documented.	Information taken from only one source and/or information not accurate. Ideas and sources not documented.

Possible Solutions	Carefully drawn, annotated and dimensioned sketches of 3 possible solutions.	Carefully drawn, sketches of 3 possible solutions. Some annotations and dimensions missing.	Carefully drawn, sketches of 2 possible solutions. Some annotations and dimensions missing.	Drawn, sketch of 1 possible solution. Some annotations and dimensions missing.
Function	Solutions appear to function well, solving the problems outlined in the design brief.	Solutions appear to function, solving most of the problems outlined in the design brief.	Structure functions pretty well, but does not solve all the problems outlined in design brief.	Fatal flaws in function. Solution does not solve many of the problems outlined in design brief.
Inventor Creation of Final Solution	Great care taken in construction process so that the creative structure is neat, attractive and follows plans accurately.	Constuction was careful and accurate for the most part, but 1-2 details could have been refined for a more useful and attractive product.	Construction accurately followed the plans, but 3-4 details could have been refined for a more useful and attractive product.	Construction appears careless or haphazard. Many details need refinement for a useable attractive product.
Engineer Notebook	Notes provide a complete record of planning, construction, testing, modifications, reasons for modifications, and some reflection about the strategies used and the results.	Journal provides a complete record of planning, construction, testing, modifications, and reasons for modifications.	Journal provides quite a bit of detail about planning, construction, testing, modifications, and reasons for modifications.	Journal provides very little detail about several aspects of the planning, construction, and testing process.