BAU Guidelines for Effective Student Assessment
We are what we repeatedly do. Excellence, therefore, is not an act but a habit.

_Aristotle_  
_Ancient Greek Philosopher and Scientist_  
_384 BC-322 BC_
President’s Message

Higher education at University level has recently undergone significant development and advancement. Such rapid progress is being driven by new directions and trends, which address the current challenges and aspirations of the contemporary academic community. Long established systems, methods and procedures are being reviewed, challenged, and enhanced. Modification and enhancement are being sought at different levels and through different tracks. One of these tracks pertains to the ways in which universities assess and evaluate the performance of students through various disciplines and different stages of studying. In the academic arena, new theories emerge and expand into advanced levels of application, reflecting the growing concern that learning objectives are being achieved, and that assessment is being carried out through valid, reliable and sound strategies and procedures.

Based on the above standpoint, Beirut Arab University issues this concise document as an attempt to provide further support and guidance to its teaching staff members. The chief aim is to offer a set of practical guidelines and principles that provide useful insight and assist examiners in formulating tests and examinations. We hope that the contents of this booklet will be of prime interest to our staff, in order to implement a high-quality testing strategy, and to adopt a continuous assessment policy employing the full spectrum of test types and formats. The guidelines for effective student assessment presented in this document represent an important and valuable contribution towards the achievement of Beirut Arab University's mission.

Prof. Dr. Amr Galal El Adawi
Editorial Committee

• Prof. Dr. Amr Galal El Adawi
  President

• Prof. Dr. Mayssah El Nayal
  Dean - Faculty of Arts

• Prof. Dr. Hassan Abdel Salam
  Dean - Faculty of Architectural Engineering

• Prof. Dr. Abdalla El Lakany
  Dean - Faculty of Pharmacy

• Mrs. Abeer Ghazal
  Technical Office

Language Editing by:
  - Prof. Dr. Olga Mattar (Chairperson-English Department)
    (Director, Language Center, BAU)

  - Dr. Najlaa Abou Agag (Ass. Professor-English Department-Faculty of Arts)

Cover page designed by:
  - Dr. Hisham El Arnaouty (Ass. Professor-Faculty of Architectural Engineering)
Contents

1. Introduction ................................................................. 6
  1.1 The Need For This Booklet .............................................. 6
  1.2 What is Assessment .................................................... 8
  1.3 Objectives of Assessment ............................................. 8

2. Typology of Student Assessment ................................. 11
  2.1 The Cognitive Domain .................................................. 11
  2.2 The Nature of Student Assessment .................................. 15
  2.3 Types of Test Items ..................................................... 16

3. Practical Considerations for Test Items ......................... 17
  3.1 Planning of Test Content ............................................. 17
  3.2 Preparing Objective Test Items ...................................... 21
  3.3 Grades and Grading ..................................................... 40

4. General Rules ................................................................. 43

5. Conclusion ........................................................................ 45

References ........................................................................... 46
1. Introduction

“If we wish to discover the truth about an educational system we must look into its assessment procedures” (Rowntree, 1977, p1)

1.1 The Need For This Booklet

Education at the University level is considered as an important foundation and catalyst for the advancement and evolvement of nations. Certainly, the progress and development of communities are measured by their achievement at the University level both in terms of quantity and quality. High quality education is also perceived as an instrument to sustain the welfare, growth and progress of communities. Furthermore, it reflects the contribution of a nation and its participation and involvement to the overall prosperity of human culture.

University education programs aim, in general, to allow individuals to learn, expand their horizons and gain various skills, as well as practical, intellectual, and professional knowledge, in order to ultimately help graduates to meet the challenges and excel in their future careers.

Learning, gaining specialized knowledge and pursuing university-level education is a dynamic, complex, and rapidly evolving domain. Central to higher education is the inherent competition among learners and the evaluation and assessment of performance among students.

Exam Design is considered as one of the most important and
influential aspects in the judgment of the learners’ achievement.

The booklet is a compilation of material, and due to limitations of space and time, it was compiled with high selectiveness. It does not claim to be comprehensive, nor does it claim to be exhaustive. While it is intended as a tool that can assist staff members in assessment procedures, it remains open for revision and welcomes additions.

The main criteria of student testing at university are based on fairness and objectivity, and the main objective of this booklet is to help achieve those.

Gaining knowledge and acquiring skills are ensured through the adequate delivery of progressive learning methods; and appropriate assessment and testing processes at each stage of study.

The success of any academic program is often measured by the extent to which the “Intended Learning Outcomes”, commonly referred to as ILOs, are achieved and maintained. Therefore, this booklet aims to highlight the chief concerns and most common difficulties that teaching staff often encounter when designing exams and formulating questions. Towards this end, an Arabic version of this document will be released for use and future reference by staff.
1.2 What is Assessment

Assessment in higher education has been defined as “The systematic collection, review and use of information about education programs undertaken for the purpose of improving student learning and development” (Marches, 1987). The overriding purpose of assessment is to understand how education programs are working and to determine whether they are contributing to student growth or development.

Assessing and/or measuring the achievement of ILOs is carried out by administering exams and tests in their various types and forms (classification of tests and their typology will be addressed later in this booklet, e.g. theoretical/practical, written/oral/interactive). The formulation of an adequate policy for evaluation and testing is thus crucial for the realization of any academic mission and objectives.

1.3 Objectives of Assessment

Concerns and worries regarding exams and testing are manifold yet common and often shared by a large section of teaching staff at university-level. They mainly include the following:

1. Disregarding to measure one or more of the skills that the learner should have acquired.

2. Limiting exams to one/few types of questions rather than including different types addressing different cognitive levels and intellectual capabilities.
3. Focusing on one part or topic in the program rather than including all the topics.

4. Allowing the learner more time for solving the different sections in an exam which does not correspond to the quantity of information provided to the learner in a certain topic.

5. Unfair distribution of grades and their irrelevance to the quantity of information.

6. Use of vague or ambiguous vocabulary in formulating questions and other test parts, potentially causing conflict in interpreting terminology, and thus confusing or misleading the respondents.

Addressing these concerns along with other worries which the teaching staff might have constitutes an important academic preoccupation, and underlies the need for a rigorous review of a university’s policy in the field of student assessment, testing and examinations. The principal aim of this review is that it provides the means by which:

• Formal learning outcomes are objectively assessed.
• Academic and professional standards are maintained.
• Credibility of the Institution’s academic awards is safeguarded.

Accordingly, there are several considerations to be carefully observed in adopting a clear policy and administering the diverse types and formats of exams.
The contents of this booklet will address the most significant issues and attempt to give relevant views and guiding principles. These will present a useful insight into the process of student assessment and help achieve the following:

1. Outlining a high quality assessment strategy.
2. Improving the teaching and learning processes.
3. Ensuring objectivity in testing and grading.
4. Improving the procedures and practices used to administer tests and exams on both the academic and management tiers.
5. Emphasizing achievement and performance rather than content and curriculum.
2. Typology of Student Assessment

2.1 The Cognitive Domain

There have been several systems and schemes proposed and developed to organize and structure the assessment of students completing their higher education at several universities. Such schemes mainly lead to establishing sets of procedures and/or practices, selectively adopted to ensure adequate evaluation and assessment through different techniques and methods of examination.

Bloom’s Taxonomy is probably the most widely employed scheme for labeling and articulating the levels of cognitive processes in test construction today. Discussions during the 1948 Convention of the American Psychological Association led Bloom – Professor Benjamin Bloom of Chicago University – to spearhead a group of educators who eventually undertook the ambitious task of classifying educational goals and objectives. Their intent was to develop a method of classification for thinking behaviors that were believed to be important in the processes of learning. Eventually, this framework became a taxonomy of three domains:

- The cognitive-knowledge based domain, consisting of six levels
- The affective-attitudinal based domain, consisting of five levels
- The psychomotor-skills based domain, consisting of six
levels. Bloom’s Taxonomy is a multi-tiered model of classifying thinking according to six cognitive levels of complexity.

- The lowest three levels are: knowledge, comprehension, and application. The highest three levels are: analysis, synthesis, and evaluation.

- The taxonomy is hierarchical; each level is subsumed by the higher levels. In other words, a student functioning at the application level has also mastered the material at the knowledge and comprehension levels.

According to Bloom’s Taxonomy, human thinking skills can be broken down into the following six categories:

- **Knowledge**: remembering or recalling appropriate, previously learned information to draw out factual (usually right or wrong) answers. It is recommended to use words and phrases such as: how many, when, where, list, define, tell, describe, identify, etc., to draw out factual answers, testing students’ recall and recognition.
• Comprehension: grasping or understanding the meaning of informational materials. It is recommended to use words such as: describe, explain, estimate, predict, identify, and differentiate, etc., to encourage students to translate, interpret, and extrapolate.

• Application: applying previously learned information (or knowledge) to new and unfamiliar situations. It is recommended to use words such as: demonstrate, apply, illustrate, show, solve, examine, classify, and experiment, etc., to encourage students to apply knowledge to situations that are new and unfamiliar.

• Analysis: breaking down information into parts, or examining (and trying to understand the organizational structure of) information. It is recommended to use words and phrases such as: what are the differences, analyze, explain, compare, separate, classify, and arrange, etc., to encourage students to break information down into parts.

• Synthesis: applying prior knowledge and skills to combine elements into a pattern not clearly there before. It is recommended to use words and phrases such as: combine, rearrange, substitute, create, design, and invent, what if, etc., to encourage students to combine elements into a pattern that’s new.

• Evaluation: judging or deciding according to some set of criteria, without real right or wrong answers. It is recommended to use words such as: assess, decide, measure, select, explain, conclude, compare, and sum-
marize, etc., to encourage students to make judgments according to a set of criteria

**Revised Bloom’s Taxonomy (RBT)**

During the 1990’s, a former student of Bloom’s, Lorin Anderson, led a new assembly which met for the purpose of updating the taxonomy, hoping to add relevance for 21st century students and teachers.

Basically, Bloom’s six major categories were changed from noun to verb forms. Additionally, the lowest level of the original, knowledge was renamed and became remembering. Finally, comprehension and synthesis was renamed into understanding and creating.

![Revised Bloom’s Taxonomy](image)

The new terms of Bloom’s Taxonomy are defined as:

- **Remembering**: Retrieving, Recognizing, and Recalling relevant knowledge from long-term memory.
- **Understanding**: Constructing meaning from oral, written, and graphic messages through interpreting, exem-
plifying, classifying, summarizing, inferring, comparing, and explaining.

- **Applying**: Carrying out or using a procedure through executing, or implementing.

- **Analyzing**: Breaking material into constituent parts, determining how the parts relate to one another and to an overall structure or purpose through differentiating, organizing, and attributing.

- **Evaluating**: Making judgments based on criteria and standards through checking and critiquing.

- **Creating**: Putting elements together to form a coherent or functional whole; reorganizing elements into a new pattern or structure through generating, planning, or producing.

### 2.2 The Nature of Student Assessment

Student assessment can either be formative or summative depending upon when the assessment is implemented and how the results are used.

- **Formative assessment** is testing conducted during the teaching of the course; it provides information useful in improving or shaping student learning. Usually, instructors implement formative types of assessment in order to help ensure that students learn and achieve course learning outcomes.
**b. Summative assessment** is testing conducted at the end of a unit or course to provide a final judgment of student performance with no chance for improvement. Summative assessments could be unit exams, final exams, or portfolios that are administrated at the end of a course.

### 2.3 Types of Test Items

**a. Objective items,**

- require students to select an answer from a set of alternatives (MCQs, True-False)
- are used with large groups of students
- are used when highly reliable test scores are needed
- are used when impartiality of evaluation is needed

**b. Subjective items,**

- permit students to express their opinion and to be original (Essay-questions, problem solving)
- are used with small groups of students
- are used when there is high interest in exploring students’ attitudes and opinions
- are only used when the teacher is confident in raters’ ability to remain fair.
3. Practical Considerations For Test Items

3.1 Planning of Test Content

Test Blueprint:
To ensure that all objectives of a given course are represented in the exam, a test blueprint could be established. A test blueprint is a two-way grid, in which the objectives of a given course, are matched with the levels of item complexity.

Table (1): A Hypothetical Test Blueprint

<table>
<thead>
<tr>
<th>Levels of items complexity</th>
<th>Remembering</th>
<th>Understanding</th>
<th>Applying</th>
<th>Analyzing</th>
<th>Evaluating</th>
<th>Creating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objec.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objec.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objec.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objec.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objec.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objec.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objec.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Test Specification:
Once the learning objectives to be included in a test are determined, a test specification should be made. That consists of a matrix, representing the number of questions within each topic and level of objectivity. It identifies the objectives and skills that are to be tested and the relative weight given
to each in the test, and ensures the desired coverage of topics and level of objectivity. Then the exam can be written.

Table(2): A Test specification matrix (a 40 item exam)

<table>
<thead>
<tr>
<th></th>
<th>Topic A</th>
<th>Topic B</th>
<th>Topic C</th>
<th>Topic D</th>
<th>Total</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remembering</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>12.5 %</td>
</tr>
<tr>
<td>Understanding</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>7</td>
<td>17.5 %</td>
</tr>
<tr>
<td>Applying</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>15</td>
<td>37.5 %</td>
</tr>
<tr>
<td>Analyzing</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>10</td>
<td>25 %</td>
</tr>
<tr>
<td>Evaluating</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
<td>2</td>
<td>5 %</td>
</tr>
<tr>
<td>Creating</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2.5 %</td>
</tr>
<tr>
<td>TOTAL</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>40</td>
<td>100%</td>
</tr>
</tbody>
</table>

The important point in the above is the idea that student knowledge is not all equal; there are levels of mastery ranging from simple recitation of facts, to formulating informed opinions regarding complex issues. An awareness of these levels can help teaching staff determine how well students really know the course content.

Once the teaching staff create the test specification they should write the items to match the level of objectivity within each topic area. To assist them throughout this, it may be helpful to refer to the following chart (Table 3), which lists Bloom’s Revised levels. The chart includes learner action at each level, question cues which can be used to develop questions, and examples of questions written to assess that level.
Table (3): Revised Bloom’s levels Chart

<table>
<thead>
<tr>
<th>Level</th>
<th>Learner Action</th>
<th>Question Cues</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remembering</td>
<td>Recall content in the exact form that it was presented. Memorization of definitions, formulas, or procedures are examples of knowledge-level functioning.</td>
<td>List, define, label, identify, name</td>
<td>Define compound interest.</td>
</tr>
<tr>
<td>Understanding</td>
<td>Restate material in their own words, or can recognize previously unseen examples of a concept.</td>
<td>Describe, associate, categorize, summarize</td>
<td>Given a list of examples, categorize the cases of compound interest versus simple interest.</td>
</tr>
<tr>
<td>Applying</td>
<td>Apply rules to a problem, without being given the rule or formula for solving the problem.</td>
<td>Apply, calculate, illustrate, solve</td>
<td>If interest on $100 is compounded daily for 14 months at 10%, calculate the total amount of interest earned?</td>
</tr>
<tr>
<td>Analyzing</td>
<td>Break complex concepts or situations down into their component parts, and analyze how the parts are related to one another.</td>
<td>Analyze, compare, separate, order, explain</td>
<td>Using the previous example, if interest were compounded monthly instead of daily, what would the difference in interest be?</td>
</tr>
<tr>
<td>Evaluating</td>
<td>Evaluating</td>
<td>Creating</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>Evaluate or make judgments on the worth of a concept, object, etc. for a purpose.</td>
<td>Assess, decide, grade, recommend, explain, judge</td>
<td>Given a list of three potential investments, including their interest rates, lengths of investment, and compounding schedule, select the best option, and defend your decision</td>
<td></td>
</tr>
<tr>
<td>Rearrange component parts to form a new whole.</td>
<td>Combine, modify, rearrange, «what-if»</td>
<td>What interest rate is required for $100 to grow to $125 in six months, compounded daily?</td>
<td></td>
</tr>
</tbody>
</table>
Time Guidelines:

Table(4): The Time Guidelines for Test Questions

<table>
<thead>
<tr>
<th>Type of Question</th>
<th>Time to Complete</th>
</tr>
</thead>
<tbody>
<tr>
<td>True-False</td>
<td>½ - 1 minute</td>
</tr>
<tr>
<td>Multiple-Choice</td>
<td>1 minute</td>
</tr>
<tr>
<td>Short-Answer</td>
<td>2 minutes</td>
</tr>
<tr>
<td>Limited Essay</td>
<td>10-15 minutes</td>
</tr>
<tr>
<td>Broad Essay</td>
<td>30 minutes</td>
</tr>
<tr>
<td>Matching exam</td>
<td>½ minute per step</td>
</tr>
<tr>
<td>Oral exam</td>
<td>5-30 minutes</td>
</tr>
</tbody>
</table>

The Staff Member is advised to:

- Allow time for distributing and collecting exams
- Allow another 5 minutes for students to review their exam
- Allow students an extra amount of time for him/her to complete the exam.

3.2 Preparing Objective Test Items

This part of the booklet presents successive sets of guidelines for preparing test items in an objective way. They are classified under seven categories of tests. For each category, relevant principles and guidelines are given in the same order. First, a broad definition is given; then special considerations followed by general guidelines and finally relevant examples.
Oral testing items:

Oral tests are becoming more common as the University procedures progress. During an oral test, both the knowledge of a subject and students’ communication skills are evaluated. Oral tests may take place in front of one person, a panel or a class.

Formal oral tests usually follow a list of questions. The examinees should not give out any extra information unless he/she has been asked to. But during informal tests, he/she can give longer responses with more information.

Intended objectives of this type of test and exam:

1. Providing direct and visual contact between examiner and examinee.
2. Assessing certain skills which are difficult to present and thus measure through other types of tests (e.g. interpretation; responsiveness; awareness; cross-topic familiarity).
3. Training respondents in oral presentation styles, consistency and objective discussion.
4. Measuring capacity to explain, describe and clarify verbally, and to orally express valid ideas and information, so as to ultimately convince referees.

• Special Considerations:

1. Allowing adequate time for each respondent.
2. Allowing equal duration for individual presentations successively.
3. Designing a panel of at least two referees.
4. Organizing the discussion process with the student (e.g. one speaker to lead the conversation; or otherwise one question per referee, etc.).
5. Sharing views about the evaluation between the panel members, while ensuring independent grading/scores on separate sheets.
6. Ensuring impartiality, and avoiding personal biases, pre-conceptions and prejudices.
7. Setting an adequate place/location with a conducive environment for oral discussions.
8. Avoiding interruption to the speaker, and keeping any questions, enquiries and comments to the end, after a student is done with his/her verbal presentation (in keeping with the time allowed for each presentation individually).

• General Guidelines:

1. Pre-planning of the test location, zoning and spatial distribution, record sheets, circulation of students and referees.
2. Pre-exam orientation and explanations to students about content, timing and best practice.
3. Pre-exam explanation to referees/examiners about the intended assessment, requisites, and general direction set by the Faculty and/or University.
4. Commitment on the part of all participants, both academic and non-academic, and organized attendance
and streaming.
5. Provision of interim breaks in case of long sessions and large numbers of students.
6. Avoiding interference which may occur if two or more students are giving their presentations at the same time to different panels.
7. Promoting and maintaining an overall positive, encouraging and enticing atmosphere throughout the test duration.

- **Oral Test Example:**

A part of a curriculum was evaluated by testing student performance through an oral test. To develop the rating scale, students were interviewed and recorded for the scorers to examine. The scorers listened to several examples and divided them into two broad categories: students who spoke well and those who did not. The scorers then discussed the rankings of these performances and created a scale based on their impressions. During the discussion of the tapes, it was decided that one of the most important features affecting fluency was the completeness of the answer followed by the number and the length of the pauses, speaking rate and finally the number of repetitions. The examples were then examined for these features which the scorers identified as influencing their perceptions of fluency. This formed the basis for the rating scale which was then used by the scorers to double mark the other performances. The rating scale designed by this process is shown in Table 5.
Table(5): The Rating Scale for the Oral Test

<table>
<thead>
<tr>
<th></th>
<th>partial answer</th>
<th>--Content--</th>
<th>complete answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>long and many</td>
<td>--Pauses--</td>
<td>short and few</td>
</tr>
<tr>
<td>2</td>
<td>slow</td>
<td>--Rate--</td>
<td>fast</td>
</tr>
<tr>
<td>3</td>
<td>many</td>
<td>--Repetitions--</td>
<td>few</td>
</tr>
</tbody>
</table>

**Performance tests:**

Performance tests (Practical testing) ask students to demonstrate proficiency in conducting an experiment, executing a series of steps in a reasonable amount of time, following instructions, creating drawings, manipulating materials or equipment, or reacting to real or simulated situations. Performance tests can be administered individually or in groups. They are logistically difficult to set up, hard to score, and the content of most courses does not necessarily lend itself to this type of testing. However, performance tests can be useful in classes that require students to demonstrate their skills.

- **Special Considerations:**
  1. Increasing testing time and the number of items per examinee.
  2. Maximizing the size of the performance assessment prompt pool to allow for adequate selection in the test assembly process.
  3. Structuring the development process to include a small scale pilot administration prior to the field test.
  4. Identifying the performance criteria at the same time
that the performance assessment items are being written.

5. Aligning the performance criteria with the curriculum standards being assessed.

• General Guidelines:

1. Specify the criteria to be used for rating or scoring (for example, the level of accuracy in performing the steps in sequence or completing the task within a specified time limit).

2. State the problem so that students know exactly what they are supposed to do (if possible, conditions of a performance test should mirror a real-life situation).

3. Give students the chance to perform the task more than once or to perform several task samples.

4. Finalize the scores after the field test papers have been evaluated.

5. Report scores at the group level rather than the individual level, to increase the dependability of the generalization.

• Following are some methods that have been used successfully to assess performance:

Open-ended or extended response exercises are questions or other prompts that require students to explore a topic orally or in writing. Students might be asked to describe their observations of a science experiment, or present arguments an historic character would make concerning a particular proposition.
For example: How would Abraham Lincoln explain causes of the Civil War?

Extended tasks are assignments that require sustained attention in a single work area and are carried out over several hours or longer. Such tasks could include drafting, reviewing, and revising an essay; conducting and explaining the results of a science experiment on photosynthesis; or even painting a car in Auto Shop.

Portfolios are selected collections of a variety of performance-based work. A portfolio might include a student’s “best pieces” and the student’s evaluation of the strengths and weaknesses of several pieces. The portfolio may also contain some “works in progress” that illustrate the improvements the student has made over time.

These methods, like all types of performance assessment, require that students actively develop their approaches to the task under defined conditions, knowing that their work will be evaluated according to agreed-upon standards. This requirement distinguishes performance assessment from other forms of testing.
Extended open-response items (Essay tests):

There are two major purposes for using essay questions that address different learning outcomes. One purpose is to assess students’ understanding of subject-matter content. The other purpose is to assess students’ writing abilities. Thus it enables the teachers to judge the students’ abilities to organize, integrate, interpret material, and to express themselves in their own words. Research indicates that students study more efficiently for essay-type examinations than for selection (multiple-choice) tests: students preparing for essay tests focus on broad issues, general concepts, and interrelationships rather than on specific details, and this studying results in somewhat better student performance regardless of the type of exam they are given. Essay tests also give the teaching staff an opportunity to comment on students’ progress, the quality of their thinking, the depth of their understanding, and the difficulties they may be having. However, because essay tests pose only a few questions, their content validity may be low. In addition, the reliability of essay tests is compromised by subjectivity or inconsistencies in grading.

- **Special Considerations:**
  1. The essay questions match the intended learning outcomes
  2. The task is specifically and clearly defined.
  3. The relative point value and the approximate time limit are specified.
• **General Guidelines:**

1. Clearly define the intended learning outcome to be assessed by the item.
2. Avoid using essay questions for intended learning outcomes that are better assessed through other kinds of assessment.
3. Define the task and shape the problem situation.
4. Specify the relative point value and the approximate time limit in clear directions.
5. State the criteria for grading
6. Use several relatively short essay questions rather than one long essay.
7. Avoid the use of optional questions
8. Improve the essay question through preview (before handing out the essay question to the students) and review (after receiving the student responses).

• **Examples:**

**Example A:** What was the full name of the man who assassinated President Abraham Lincoln?

**Example B:** State the full name of the man who assassinated President Abraham Lincoln and explain why he committed the murder.

There is just one single correct answer to Example A because the students need to write the full name of the man who assassinated President Abraham Lincoln. The question assesses verbatim recall or memory and not the ability to think. For this reason, Example A would not be considered a typical
essay question; it is a kind of short-answer question. Example B assesses students’ understanding of the assassination and it is more effective at providing students with the opportunity to think and to give a variety of answers. Answers to this question may vary in length, structure, etc.

**Short-answer tests:**

Short answer questions usually ask students to list, name, define, or identify. Depending on the objectives, short-answer questions can call for one or two sentences. Short-answer tests are easier to write, though they take longer to score, than multiple-choice tests. They also give the teaching staff some opportunity to see how well students can express their thoughts, though they are not as useful as longer essay responses for this purpose.

- **Special considerations:**
  1. Use the short-answer question when the aim is for the examinees to recall the answer than merely select it
  2. Clarify the type of information requested.
  3. Provide sufficient time for answering.

- **General Guidelines:**
  1. Use paraphrases rather than literal textbook statements in posing questions.
2. If a specific number of points is desired in an answer, state that number in the question.

3. Specify the terms in which answers should be given.

- **Here is an example of the short-answer question type:**

  The thin membrane that separates the inner ear from the external ear is commonly called the ____________.

  This question is very clear in its desired response. In addition, it assesses recall of knowledge-level processing.

  Short-answer question types can also include asking for definitions and short lists. For example:
  1. Briefly define insectivore.
  2. List the three states that comprise the west coast of the continental United States.

**Matching tests:**

The matching format is an effective way to test students’ recognition of the relationships between words and definitions, events and dates, categories, examples, and so on. Matching items are not conceptually difficult. They consist of a set of associated responses, and some directions for drawing the most appropriate correspondence between them. These items are especially useful when a large number of specific facts are expected to be recognized.
• **Special considerations:**

  1. Matching items require explicit directions in order not to confuse the examinees about how or where to respond.

  2. The wording must be clear and concise.


• **General Guidelines:**

Matching items can assess a large amount of information in a confined space on the exam page, relative to multiple-choice questions. If developed carefully, the probability of guessing is low. To decrease that probability further, avoid equal-sized lists by including a few “distracter” items in the second (answer) column.

• **Matching Test Example:**

  Directions: On the line next to each author in Column A, place the letter of the genre in Column B for which the author is best known.

  Answers in Column B may be used only once.

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
</tr>
</thead>
<tbody>
<tr>
<td>_____ 1. James Michener</td>
<td>A. History</td>
</tr>
<tr>
<td>_____ 2. Stephen King</td>
<td>B. Horror</td>
</tr>
<tr>
<td>_____ 3. Erma Bombeck</td>
<td>C. Humor</td>
</tr>
<tr>
<td>_____ 4. Agatha Christie</td>
<td>D. Mystery</td>
</tr>
<tr>
<td>_____ 5. Walt Whitman</td>
<td>E. Poetry</td>
</tr>
<tr>
<td>_____ 6. Danielle Steele</td>
<td>F. Romance</td>
</tr>
<tr>
<td>_____ 7. Isaac Asimov</td>
<td>G. Science Fiction</td>
</tr>
</tbody>
</table>
The following example is considered a well-written matching question:

Match the function of the computer part in Column A with its name in Column B:

1. Collection of transistors used to manipulate data  
   a. BIOS

2. Uses laser beam to read spirals of indentations  
   b. CMOS

3. Serves as intermediary between OS and hardware components  
   c. Processor

4. Translates image information into electrical currents  
   d. Video card

   e. CD-ROM

   f. Serial port

Multiple-choice tests:

This is the most popular form of objective testing. Not only are multiple-choice tests relatively easy to correct, but they can also be designed so that they measure a variety of learning tasks.

Multiple-choice items can be used to measure both simple knowledge and complex concepts. Since multiple-choice questions can be answered quickly, a staff member can assess the students’ mastery of many topics in a one-hour exam. In addition, the items can be easily and reliably scored.
An MCQ exam requires a certain degree of perfection in order to achieve its goals. Items of this type include a stem followed by a number of options from which the student indicates \textbf{ONLY ONE} option. The others are distracter options.

- **Special Considerations:**

1. Make sure the item can be answered without looking at the options OR that the options are 100% true or false.

2. Include as much of the item as possible in the stem; the stems should be long and the options short.

3. Avoid superfluous information.

4. Avoid “tricky” and overly complex items.

5. Write options that are grammatically consistent and logically compatible with the stem; list them in logical or alphabetical order.

6. Write distracters that are plausible and are as long as the other options.

7. Avoid using absolutes such as: always, never, and all in the options; also avoid using vague terms such as: usually and frequently.

8. Avoid negatively phrased items (i.e., those with except or not in the lead-in). If it is necessary to use a negative stem, use only short (preferably single word) options.
9. The number of options is from six to four and most important of all focus on important concepts... Don’t waste time testing trivial facts.

- **General Guidelines that Tip Off Students to the Correct Answer:**

  1. Grammatical cues - one or more distracters don’t follow grammatically from the stem.
  2. Logical cues - a subset of the options is collectively exhaustive.
  3. Absolute terms - terms such as “always” or “never” are in some options.
  4. Long correct answer - correct answer is longer, more specific, or more complete than other options.
  5. Word repeats: a word or phrase is included in the stem and in the correct answer.
  6. Convergence strategy: the correct answer includes the most elements in common with the other options.

- **General Guidelines that Make Items Unnecessarily Difficult:**

  1. Options are long or complicated, or contain duplicated portions.
  2. Numeric data are not stated consistently.
  3. Terms in the options are vague (examples--”rarely,” “usually”).
  4. Language in the options is not parallel.
  5. Options are in an illogical order.
6. “None of the above” is used as an option.
7. Stems are tricky or unnecessarily complicated.
8. The answer to an item is “hinged” to the answer of a related item.

- **Example of Multiple-Choice Questions:**

Question -1-
Was the infantry invasion of Japan a viable alternative to the use of the atomic bomb to end World War II? If so, why? If not, why not?
A. Yes; transport ships were available in sufficient numbers.
B. Yes; island defenses in Japan were minimal.
C. No; estimated casualties would have been much greater.*
D. No; Japan was on the verge of having an atomic bomb.
* Wanted answer.

**True-false tests:**

True/false items are also popular in objective examinations. Unlike multiple-choice questions, there are only two options. Details tend to be emphasized in this type of test, so students need to note details carefully while studying. True-False items can provide:
- the widest sampling of content or objectives per unit of testing time.
- scoring efficiency and accuracy.
- versatility in measuring all levels of cognitive ability.
- highly reliable test scores.
- an objective measurement of student achievement or ability.

• **Special Considerations:**

**This type of test:**
1. Incorporates an extremely high guessing factor. For simple true-false items, each student has a 50/50 chance of correctly answering the item without any knowledge of the item’s content.
2. Can often lead an instructor to write ambiguous statements due to the difficulty of writing statements which are unequivocally true or false.
3. Does not discriminate between students of varying ability as good as other item types.
4. Can often include more irrelevant clues than do other item types.
5. Can often lead an instructor to favor testing of trivial knowledge.

• **General Guidelines:**
1. Base true-false items upon statements that are absolutely true or false, without qualifications or exceptions.
2. Express a single idea in each test item.
3. Include enough background information and qualifications so that the ability to respond correctly to the item does not depend on some special, uncommon knowledge.
4. Avoid lifting statements from the text, lecture or other materials so that memory alone will not permit a correct
answer.
5. Avoid using negatively stated item statements.
6. Avoid the use of unfamiliar vocabulary.

**Example of True-false tests:**

True or False?

1. Everybody has a “biological clock.”
   ...TRUE ...FALSE

2. Drinking coffee cures drowsiness while driving.
   ...TRUE ...FALSE

3. I can tell when I’m going to fall asleep.
   TRUE FALSE

4. I’m a safe driver so it doesn’t matter if I’m sleepy.
   ...TRUE ...FALSE

5. I can’t take naps.
   ...TRUE ...FALSE

6. Nearly everyone gets enough sleep.
   ...TRUE ...FALSE

7. Being sleepy makes you misperceive things.
   ...TRUE ...FALSE

8. Young people need less sleep.
   ...TRUE ...FALSE
Table (6): Summary of Different Kinds of Test/Question

<table>
<thead>
<tr>
<th>Type of Test/Question Format</th>
<th>Information About the Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance Tests</strong></td>
<td>These can be administered individually or in small groups. They offer the student the opportunity to demonstrate their practical skills. Students should be informed ahead of time the scoring criteria, exactly what they are supposed to do and should be given the chance to perform a task more than once or perform several tasks.</td>
</tr>
<tr>
<td><strong>Essay Tests</strong></td>
<td>Essays offer the instructor the ability to ascertain the student’s ability to organize, integrate, and interpret material. Research has shown that students study more for essay exams. Reliability in grading can be challenged when using essay exams.</td>
</tr>
<tr>
<td><strong>Short-Answer Tests</strong></td>
<td>Students answer questions in one or two sentences or a short paragraph. This type of tests may be easier to write, but takes longer to score than MC or T/F and again reliability can be questioned in the scoring of the answers. These may give you a limited insight into how students express themselves in a written format related to a specific concept.</td>
</tr>
<tr>
<td><strong>Matching Tests</strong></td>
<td>This type of test can effectively examine a student’s ability to recognize relationships between concepts, links, etc...</td>
</tr>
<tr>
<td><strong>Multiple-Choice Tests</strong></td>
<td>These items can be designed to measure simple and complex concepts. They can also be scored quickly and with a good level of reliability.</td>
</tr>
<tr>
<td><strong>True-False Tests</strong></td>
<td>These tests are less reliable than other forms. They may be testing the students guessing ability rather than their knowledge of a subject. Some faculties include an EXPLAIN section to a T/F exam, and thus the student chooses an answer and then must explain his/her choice.</td>
</tr>
</tbody>
</table>
3.3 Grades and Grading:

A grade is a formal certification of competence and achievement of learning outcomes that should reflect as accurately as possible a student’s performance in a module or in its elements.

A valid grading system should meet three criteria:
1. It should be fair.
2. It should be clear to students so that they are able to chart their progress.
3. It should accurately reflect differences in student performance.

Awarding the grades is a matter of academic judgment against agreed criteria (learning outcomes and grade descriptors) and should not be simply a mathematical exercise. It is a way for staff members to communicate what they view as important for a student to learn and this should be backed up by what they teach. Clear grade descriptors will assist in this effort.

Grading Strategies:

1. Grade on the basis of students’ mastery of knowledge and skills. It is advisable to restrict evaluations to academic performance. Eliminate other considerations, such as classroom behavior, effort, classroom participation, attendance, punctuality, attitude, personality traits, or student interest in the course material, as the basis of course grades. If these non-academic factors are taken into consideration, the primary
objective of the grade is obscured, as an indicator of what students have learned.

2. Normative grading produces undesirable consequences for many students, such as reduced motivation to learn, debilitating evaluation anxiety, decreased ability to use feedback to improve learning, and poor social relationships.

3. Try not to overemphasize grades. Explain to the students the meaning of and basis for grades and the procedures that will be used in grading. At the beginning of the term, inform students, in writing, how much tests, papers, homework, and the final exam will count toward their final grade. Once grading policies are explained, avoid stressing grades or excessive talk about grades. This only increases students’ anxieties and decreases their motivation to do something for its own sake rather than to obtain an external reward such as a grade.

4. Keep students informed of their progress throughout the term. For each paper, assignment, midterm, or project that you grade, give students a sense of what their score means. Try to give a point total rather than a letter grade. Letter grades tend to have emotional associations that point totals lack. Do show the range and distribution of point scores, and indicate what level of performance is satisfactory. Such information can motivate students to improve if they are doing poorly or to maintain their performance if they are doing well. By keeping students informed throughout the term, unpleasant surprises at the end are prevented.
Constructive Grading:

1. Encouraging comments on a test or paper convey respect for what the student attempted to accomplish, and praise for what they did accomplish.

2. Effective grading acknowledges and reinforces the strengths of students’ work prompts students to recognize shortcomings and options and provides suggestions for improving performance.

3. Effective grading also provides feedback that helps students:

   • understand what “works” in their papers and essay tests
   • learn from their “mistakes”
   • position themselves to stretch further towards becoming more effective in the future.
4. General Rules

With regard to the formulation of written tests, the following are general rules addressing the “content” to be taken into consideration by the examiners. Constructing test items for standardized tests of achievement, ability, and aptitude is a task of enormous importance because test items are the foundation of written tests of mental attributes, and the ideas they express must be articulated precisely and succinctly:

- Use clear and simple language.
- Avoid verbosity and irrelevant information.
- Present a definite, explicit and singular question or problem in the rubric.
- Make the alternative mutually exclusive, in case multiple choice answers are used.
- Make sure there is only one correct or best answer.
- Avoid using confusing or misleading structures. To ensure continuous monitoring and feedback, staff members should also consider the following practices:
  - To publicize and implement the principles and procedures
  - To adopt processes of testing that are explicit, valid, and reliable.
  - To ensure that testing is Fair, Secure, and Rigorous.
  - To check that the amount and timing of testing enables effective and appropriate measurement of student achievements of ILOs.
  - To encourage assessment practice that promotes effective learning.
Examples to carry out this assessment include:

1. Feedback loop: Students can apply formative feedback, from staff or peer, to improve their performance in the next assessment

2. Extended assignment: Students give a topic based on their research

3. Peer assessed activities: Students in pairs or groups discuss and comment constructively on one another’s work during formal lectures.

From a complementary standpoint, and also for the purpose of effective examination and testing, further considerations should be addressed regarding “Procedural Rules”. These are however, beyond the scope of this booklet, yet remain an integral part of the testing and assessment process, and thus will require more awareness and recognition by the parties involved. They could be classified under the following six topics:

1. Duration, Timing, and Scheduling
2. Organization and Operation
3. Place, Location, and Environment
4. Supervision and Administrative Procedures
5. Marking and Reporting
6. Feedback and Right of Appeal
5. Conclusion

To conclude, there is no doubt that sustaining high-quality education at university level is considered an important strategy conducive to the development and progress of individuals and communities. Nations aiming at implementing programs of sustainable development, focus first and foremost on improving the quality and quantity of education offered, in particular to university students. The most significant and most commonly used instrument for such measurement is the assessment of the learning process. This booklet focused on assessment in the form of exams and tests administered to undergraduate students.
References


Internet References

(10) http://www.coe.uga.edu/epltt/bloom.htm

(11) http://www.hcc.hawaii.edu/intranet/committees/FacDevCom/guidebk/teachtip/quizzes.htm

(12) http://web.odu.edu/educ/llschult/blooms_taxonomy.htm


(14) http://www.utexas.edu/academic/diia/assessment/iar/students/plan/types.php

(15) http://honolulu.hawaii.edu/intranet/committees/FacDevCom/guidebk/teachtip/grading.htm

(16) http://www.iupui.edu/~profdev/handbook/chap1.html


(18) http://honolulu.hawaii.edu/intranet/committees/FacDevCom/guidebk/teachtip/quizzes.htm
To contact us visit http://iconnect.bau.edu.lb/