TEST
DEVELOPMENT
AND EVALUATION

Course Code 6462

B.Ed 4 Year

Department of secondary Teacher Education
Faculty of Education
ALLAMA IQBAL OPEN UNIVERSITY ISLAMABAD
COURSE TEAM

Chairman Course Team: Dr. Naveed Sultana
Department of Secondary Teacher Education

Course Development Coordinator: Dr. Muhammad Zafar Iqbal Mir
Dr. Muhammad Tanveer Afzal

Compiled By:
Dr. Munazza Ambreen
Dr. Muhammad Tanveer Afzal
Dr. Jahan Ara Shams
Dr. Muhammad Azeem
Safia Janjua
Hassan Arslan Khalid
Inam Jahingir
Kishwar Janjua
Shahida Parveen

Reviewers:
Dr. Naveed Sultana
Dr. Muhammad Tanveer Afzal
Dr. Muhammad Zafar Iqbal Mir

Editor: Fazal Karim

Course Coordinator: Dr. Muhammad Zafar Iqbal Mir

Formatted by: Muhammad Javed
FORWARD

Learning is an enduring change in the mechanisms of behavior. Over the last thirty years or so learning has become one of the most used words in the field of education. Adult education became lifelong learning; students became learners, teachers became facilitators of learning; schools are now learning environments; learning outcomes are carefully monitored. The message is to get the instructional regime, as learning is to measure by tests and assessments.

This course is designed for fulfilling the content expertise of prospective teachers who are enrolled in B.Ed 4 Years in Allama Iqbal Open University. This will help the teachers and students to have the knowledge of learning outcomes, how those can be measured and assess in order to make the teaching learning process functional. What should be the proper implication of tests and those can be designed properly by the prospective teachers for the future implementation in the classroom setting.

For the optimization of the student learning, it is mandatory that teacher can develop, administer, score and report the test score. Teacher must be able to enhance the validity and reliability of tests which are using in the classroom setting in order to judge students’ performance. The experience towards the development of tests and evaluation may contribute to towards the professional development of prospective teachers.

This course intends towards the professional development of the students in the field of developing tests and evaluations. In the end, I am happy to extend my gratitude to the course team chairperson, course development coordinator, unit writers and reviewers for the development of this course. Any suggestion for the improvement in this course will be warmly welcomed.

Vice-Chancellor
February, 2020
PREFACE

Learning takes place in students’ heads where it is invisible to others. This means that learning must be assessed through performance: what students can do with their learning. Assessing students’ performance can involve assessments that are formal or informal, high- or low-stakes, anonymous or public, individual or collective. In this book we provide strategies for assessing student learning and performance as well as ways to clarify to prospective teachers about the performance criteria of students.

In this course, assessment techniques are designed to gauge the effectiveness of the teaching learning process. There is a need to explain assessment standards clearly and several subsequent points when you introduce the course to your prospective teachers. This book will help them to guide all activities related to testing, their development and evaluation.

This particular course is designed to enhance the knowledge and skill of the prospective teachers towards the development of tests and evaluation of the students. The first unit is regarding the need, scope and significance of the evaluating process, second unit explains the kinds of the tests, third to sixth units are related to the development of different tests items and techniques, eighth unit includes the assembling, administration and appraising of test and the last unit has models of evaluation and their implementation.

These units are arranged and made to fulfill the requirement of the prescribed course that include the related exercises that will also help the prospective teachers to develop the attitude and skills towards the evaluation process of the students.

In the end, I am thankful to the course team and appreciate their efforts for the development of this course.

Dr. Naveed Sultana
Chairperson STED
COURSE INTRODUCTION

This course is designed to help prospective teachers to measure, test and evaluate their students’ performance. As without measuring or assessing the performance of the student, learning cannot be advanced. The tests give a path to ensure an entity of knowledge which is being inculcated by the teacher into the students. So a teacher should have a sound knowledge of conducting tests, use of measurements and evaluation in gaining insight of teaching learning process.

Measurement, testing and evaluation are the important terms associated with the realm of education. Providing education without testing and measurement may not be fruitful to achieve the educational end, therefore the educators and the researchers are focusing more on the testing and the evaluation. This particular course aim to the development of the measurement tools particularly the tests, both classroom test and the high-stake testing has been included. The course also aims at evaluation of the students’ progress activities assume that the prospective teachers may have in depth understanding of the evaluation process after studying different models of evaluation.
OBJECTIVES

At the end of the course the students will be able to:
• use the concept of evaluation for monitoring student’s progress.
• highlight the importance of high stake and classroom testing
• use different kinds of test to monitor the students’ progress.
• develop the supply type and selection type test items for the use of educational evaluation.
• develop section type test items for the use of educational evaluation.
• use the different models and the techniques of evolution for course, program and the institutional evaluation.
## CONTENTS

<table>
<thead>
<tr>
<th>Unit#</th>
<th>Topics</th>
<th>Page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit-1:</td>
<td>Evaluating Students Progress: Need and Significance</td>
<td>01</td>
</tr>
<tr>
<td>Unit-2:</td>
<td>Need, Scope and Kinds of Tests</td>
<td>23</td>
</tr>
<tr>
<td>Unit-3:</td>
<td>Classroom Testing and High Stake Testing</td>
<td>39</td>
</tr>
<tr>
<td>Unit-4:</td>
<td>Development of Supply Test Items</td>
<td>69</td>
</tr>
<tr>
<td>Unit-5:</td>
<td>Development of Selection Type Test Items</td>
<td>97</td>
</tr>
<tr>
<td>Unit-6:</td>
<td>Developing Alternate Techniques – I</td>
<td>111</td>
</tr>
<tr>
<td>Unit-7:</td>
<td>Developing Alternate Techniques – II</td>
<td>135</td>
</tr>
<tr>
<td>Unit-8:</td>
<td>Assembling, Administering and Appraising Tests</td>
<td>159</td>
</tr>
<tr>
<td>Unit-9:</td>
<td>Different Models of Evaluation and Their Implications</td>
<td>181</td>
</tr>
</tbody>
</table>
EVALUATING STUDENTS PROGRESS: NEED AND SIGNIFICANCE

Written By: Dr. Muhammad Tanveer Afzal
Reviewed By: Dr. Naveed Sultana
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>03</td>
</tr>
<tr>
<td>Objectives</td>
<td>04</td>
</tr>
<tr>
<td>1.1 Concept of Evaluation</td>
<td>05</td>
</tr>
<tr>
<td>1.2 Use of Evaluation For Different Purposes: Student and Teacher Evaluations</td>
<td>06</td>
</tr>
<tr>
<td>1.3 Measuring Students’ Progress: Basic Assumption</td>
<td>08</td>
</tr>
<tr>
<td>1.4 Need for Evaluating Students’ Progress</td>
<td>10</td>
</tr>
<tr>
<td>1.5 Accountability and Evaluation</td>
<td>13</td>
</tr>
<tr>
<td>1.6 Selection of Appropriate Tools for Students Evaluation</td>
<td>14</td>
</tr>
<tr>
<td>1.7 Use of Evaluation for Instructional Feedback</td>
<td>15</td>
</tr>
<tr>
<td>1.8 Self Assessment Questions</td>
<td>20</td>
</tr>
<tr>
<td>Bibliography</td>
<td>22</td>
</tr>
</tbody>
</table>
INTRODUCTION

Educational evaluation started off as a branch of psychology in the late 50s, as a result of curriculum innovations. It was then referred to as Educational Measurement, Measurement and Evaluation or Test and Measurement. Adequate educational evaluation acts as an incentive to pupils' studies. The test, examination and evaluation devices stimulate pupils' interest and enable them to make, greater efforts. Without the use of evaluation devices, most pupils will not take their learning seriously. Educational evaluation provides pupils record of attainment which can be used for selection for further education, for placement into class or job, and for guidance and counseling purposes.

The classroom teacher or evaluator should always be perfectly clear in his mind about what he is aiming to achieve i.e. what to evaluate and how to evaluate. Evaluation of educational programs should be comprehensive i.e. assess pupils' progress in all areas. Educational evaluation, apart from testing knowledge (Memorization), should also bring about pupils originality and use of ideas and their ability to think and apply the knowledge and skills already learnt.

It cannot be denied that the evaluation of teaching and learning is an exceedingly complex activity. However, the efficiency of the teacher and the growth and achievement of the pupil can be evaluated through the use of such devices as checklists, rating scales, and tests of different aspects of teaching ability, interview and questionnaires. Through the use of such devices much valuable data may be gathered relative to many of the important aspects of teaching and learning. Evaluation is perhaps the most complex and least understood of the terms. Inherent in the idea of evaluation is "value." When we evaluate, what we are doing is engaging in some process that is designed to provide information that will help us make a judgment about a given situation. Generally, any evaluation process requires information about the situation in question. A situation is an umbrella term that takes into account such ideas as objectives, goals, standards, procedures, and so on. When we evaluate, we are saying that the process will yield information regarding the worthiness, appropriateness, goodness, validity, legality, etc., of something for which a reliable measurement or assessment has been made.
OBJECTIVES

After studying this unit, the students will be able to:
1. define and describe evaluation.
2. describe the concept of evaluation with examples.
3. list the uses of evaluation for different purposes and talk about its effectiveness.
4. discuss different methods of measuring students’ progress.
5. analyze the need for evaluating students’ progress in school perspective.
6. explain the tools for students’ evaluation in school.
7. discuss use of evaluation for instructional feedback in school.
1.1 CONCEPT OF EVALUATION

Evaluation is an integral component of all systems of education at all processes. It is what enables educators, teachers, administrators, policy makers and the community have an idea of what is missing and what is available. Evaluation can be defined in two ways, depending on what we want to achieve at the end of an exercise. Evaluation is the systematic process of collecting and analyzing data in order to determine whether, and to what degree objectives have or are being achieved. Evaluation is the systematic process of collecting and analyzing data in order to make decisions. The first part of these definitions (systematic process of collecting and analyzing data) is common to both of the definitions provided here. However, the basic difference between the two definitions is the issue of whether decisions or judgments are an integral component of evaluation. The second definition seems to be more inclusive, as it does not preclude the activities implied in the first definition.

It is "Value judgment for an object or its meaning." Education evaluation is analysis and judgment of the value of an educational input, process and outcome. Evaluation is perhaps the most complex and least understood of the terms. Generally, any evaluation process requires information about the situation in question. A situation is an umbrella term that takes into account such ideas as objectives, goals, standards, procedures and so on. Some definitions of evaluation to help develop the understanding are given below.

1. Tyler (1951) "Education evaluation is the judgment process for the educational goal (behavioral objectives) realized through education and class activities."

2. Cronbach (1984) "Education evaluation is the process of information gathering and treatment necessary to make a decision for an education program."


4. The UNESCO definition of evaluation is the general process of a systematic and critical analysis leading to judgments and/or recommendations regarding the quality of a higher education institution or a programme. Evaluation is carried out through internal or external procedures. In the United Kingdom, evaluation is also called review.

5. The Glossary to the Berlin Summit of 2003 states “Apart from accreditation, evaluation is the central activity to assure quality in higher education. To evaluate means to assess teaching and academic studies in a subject or department and the related degree programmes. Strengths and weaknesses of education and training should be demonstrated by stock taking and..."
analysis and proposals should be formulated to promote its quality as well. Evaluation is carried out through internal or external procedures.”

When we evaluate, we are saying that the process will yield information regarding the worthiness, appropriateness, goodness, validity, legality, etc., of something for which a reliable measurement or assessment has been made. For example, I often ask my students if they wanted to determine the temperature of the classroom they would need to get a thermometer and take several readings at different spots, and perhaps average the readings. That is simple measuring. The average temperature tells us nothing about whether or not it is appropriate for learning. In order to do that, students would have to be polled in some reliable and valid way. That polling process is what evaluation is all about. A classroom average temperature of 75 degrees is simply information. It is the context of the temperature for a particular purpose that provides the criteria for evaluation. A temperature of 75 degrees may not be very good for some students, while for others, it is ideal for learning. We evaluate every day. Teachers, in particular, are constantly evaluating students and such evaluations are usually done in the context of comparisons between what was intended (learning, progress, behavior) and what was obtained.

1.2 USE OF EVALUATION FOR DIFFERENT PURPOSES: STUDENTS & TEACHERS EVALUATION

The main reason teachers evaluate is to find out what students have learned the outcome of the instruction. This information is used in two ways: first to inform the teachers about their teaching and what needs to be taught next and second, to make a judgment about how well students have learned the knowledge or skill being taught. Evaluation is a systematic process that involves a variety of activities. Teachers gather information about student achievement informally and formally. Informal evaluation is used by the teacher to provide feedback to students and to check for understanding in the teaching and learning process. Informal evaluation activities include observation of students as they work in groups, pretests, short classroom assignments, practice tasks, oral questioning and discussion. Formal evaluation is used to judge student achievement; that is, how well the student has learned the knowledge and/or skills. Students are marked on formal evaluation tasks and this mark is usually part of their report card grade. These evaluation tasks can include projects, writing assignments, performances, tests, reports and research. Student report card marks are based on a number of different evaluation activities over an extended time. Teachers are careful to use many opportunities to evaluate students before they make a judgment about a student's achievements.
Planning, Structuring and Conducting Instruction

Many of the decisions made by the teachers focus on planning, structuring and conducting instructional activities as instructions make the core of classroom activities. Planning and structuring decisions are made for future instructional activities, whereas, actual teaching process in classroom also requires on the spot assessment and decision making. Hence, a large proportion of teacher assessment is concerned with planning and delivering instructions.

Diagnosing Pupil’s Problems

Diagnostic decisions are made about individual students as well as about group strengths, weaknesses and needs. Typically information is gathered that will allow the teachers to diagnose the specific area that needs further attention or where progress is being made. The diagnosis includes an assessment of why a student may be having difficulty that appropriate instructional activities can be prescribed. For example, teachers use homework diagnostically to determine the extent of student’s understanding and to identify students who do not understand assignment. A pretest may be used to diagnose specific gaps in student’s knowledge that need to be targeted. Students are closely monitored to check motivation, understanding and progress. Such assessments are done in the beginning of the year to obtain an idea of the abilities and interests of the students. Much of the assessment data that teacher collect is used to identify, understand and remediate pupil’s problems and learning difficulties.

Grading Academic Learning and progress

Teachers spend much of their time on collecting information that have to be used to grade pupils or make final judgment about their academic progress. The collection of information for making final decision and assigning grades at the end of the instruction is termed as summative assessment. Because of summative evaluation teachers evaluate the effectiveness of his / her teaching and quality of learning that has taken place in the learner.

Evaluation Stimulates Students to Study

Testing serves as a stimulus to daily preparation. A questioning teacher creates incentives for students to learn more. He sets up effective and definite goals for learning; giving oral or written examination is a good incentive for the students to study harder or to do better work. It makes the learner familiar with his own results. Likewise, he needs to understand his own high and low potential for
learning, but even more, he needs help in understanding the personal problems of human relations. The teacher, by giving an unannounced or announced test of some sort will, no doubt, stimulate the pupils to study the lesson assigned or the work covered. The final examination given in high school in college at the end of the course or term furnishes a very powerful stimulus to review.

**Evaluation Helps Parents to Understand Pupil-growth, Interest and Potentialities:**

The major responsibility of the teacher is to help the parents understand their child about his progress in the various areas of the curriculum, his desires and motives and behaviour his potentialities for learning as well as his achievement.

**Evaluation Aids in Devising more Effective Instructional Materials and Procedures of Instruction**

Evaluation is also helpful in measuring the validity and reliability of instruction. The teacher should know how to measure the results of his work in order to adapt his procedure to need of the varying situations from the point of view of the measurement of teaching effectiveness finds its greatest value in the possibilities if offers for the improvement of teaching.

**Evaluation Sets up Standards of Performance for the Pupils**

It increases the effectiveness of education by setting up standards of achievement in terms of varying capacities. A standard test can be used in comparing the merits of different schools, different classrooms method, different organizations of materials and the different lengths and methods of assignment.

### 1.3 MEASURING STUDENTS PROGRESS: BASIC ASSUMPTION

One of the most challenging steps in the teaching responsibilities is the evaluation of student progress. Planning for student evaluation is an integral part of teaching, not just the final step of the instructional process. Formative evaluation can communicate to both teachers and students whether or not course content is effectively being communicated and learned, information that can lead to refinement of instruction on the part of the teacher and refinement of studying techniques on the part of the students. An evaluation system of students’ academic performance is necessary to help ensure that all students are succeeding within the framework of the educational goals and objectives. The student’s evaluation on a regular basis promotes continuous assessment of a student's
performance; informs the student, his or her parents or guardians, and the school counselor about the student's performance and progress; and provides a system of notice that allows intervention strategies to be implemented if necessary to improve the student's performance.

Meaningful evaluation includes consideration of all activities that have occurred during the particular evaluation period. Such activities should include homework, projects, reports, class participation, portfolio preparation and tests, including unit tests. The relative value attached to any activity shall be determined by the importance of the activity toward achieving the course objectives. Teachers evaluate student performance and keep accurate records in order to substantiate a grade or assessment given in a course. Academic grades are awarded to students in a manner which describes, as accurately as possible, the extent to which student growth has occurred in relationship to the requirements of curriculum content. Grading is based on the progress the individual student has made. Various evaluative techniques may be employed to assess student progress and to form the basis for determining the grade to be awarded. The determination of the grade to be awarded is a responsibility of the teacher.

Broadly stated, evaluation should be planned and implemented in ways that measure outcomes and much more with a view to enhancing the quality of intervention efforts and the long-term benefits for students and society. One purpose of evaluation is to provide feedback on efficacy so processes can be revised and fine-tuned. Such formative evaluation also includes information on participants, approaches, resources, implementation strategies, program organization, staffing, operational policies and practices. It also should include data on the characteristics of the system's "clients" -- who they are, what they want and need, how they differ from those in other locales -- as a prerequisite for effective planning and as another basis for interpreting the appropriateness of observed processes and outcomes. In addition to being helpful as a planning tool, assessment is the means by which we gauge progress. We need to have a clear sense of what our students have achieved so that we can make decisions about how we approach future instruction, where individuals need particular support and, more generally, how we are succeeding. One standout trait of teachers who make major progress with their students is frequent assessment. The focus becomes not “what I taught” but rather “what my students learned.” In this sense, assessment is vital to our mission as an organization. In order to close the achievement gap, we need to make dramatic, measurable gains with our students. We will not know if we have accomplished our goals if we do not record where our students started – and where and by how much they have grown. Assessment holds us accountable to our goals.
1.4 NEED FOR EVALUATING STUDENT’S PROGRESS

In today's education climate, school success is defined as ensuring achievement for every student. To reach this goal, educators need tools to help them identify students who are at risk academically and adjust instructional strategies to better meet these students' needs.

How Student Progress Evaluation Improves Instruction?

Student progress evaluation is a practice that helps teachers use student performance data to continually evaluate the effectiveness of their teaching and make more informed instructional decisions. Through student’s progress evaluation, the teacher determines a student's current performance level on skills that the student will be learning that school year, identifies achievement goals that the student needs to reach by the end of the year, and establishes the rate of progress the student must make to meet those goals. The teacher then measures the student's academic progress regularly, weekly, biweekly or monthly.

By regularly measuring all skills to be learned, teachers can graph changes and compare a student's progress to the rate of improvement needed to meet end-of-year goals. If the rate at which a particular student is learning seems insufficient, the teacher can adjust instruction. After noting the pattern of progress, the teacher can adjust instruction to improve student learning. If the student's performance falls below the line, the teacher may use more intense instruction (in small groups or one-on-one), re-teach the material, or provide additional opportunities for the student to practice certain skills. Research has demonstrated that when teachers monitor student’s progress, students learn more, teacher decision making improves and students become more aware of their own performance. A significant body of research conducted over the past 30 years has shown this method to be a reliable and valid predictor of subsequent performance on a variety of outcome measures and thus useful for a wide range of instructional decisions.

Why test? A Rationale for Evaluating Learning and Instruction

Assessment can help determine, if objectives were achieved and assist in the development and implementation of individualized education programs (IEPs). In addition, through evaluation, teachers can determine the direction of future instruction and develop a basis for extra help where needed. The paramount
purpose for all assessment is to gather information to facilitate decision making (Witt, Elliott, Kramer & Gresham, 1998). These may be global decisions, such as how well the student does when compared with the rest of his or her class, or local decisions, such as the material that the individual student has mastered and the material that he or she needs to review. If we think about assessment as assessment for intervention, the basic purpose is to identify changes that are needed in behaviors or environments and to decide how to accomplish the goals of the needed changes (Barnett et al., 1997).

**How Can I Integrate Instruction and Assessment?**

Teaching in an inclusive learning environment has made our teacher contributors attentive to the relationship of instruction and assessment. Assessment is seen as part of the learning process (Margaret Jenkins) and is in itself a learning experience (Karen Willig). Assessing learning is grounded in learning. Cullen and Pratt (1992) contended that continual evaluation of student learning is an integral part of the teaching and learning process and forms the basis for immediate action. Progress evaluation can help the children learn more and learn faster and help the teachers make better decisions about the type of instruction that will work best with your child. Children's progress is being monitored constantly at school, through the steady stream of homework assignments, quizzes, tests, projects and standardized tests.

**Student’s Progress Monitoring**

Student progress monitoring helps teachers evaluate how effective their instruction is, either for individual students or for the entire class. If the child is meeting or exceeding the expectation, the teacher continues to teach the child in the same way. If the child's performance on the measurement does not meet the expectation, then the teacher changes the teaching. The teacher might change the method being used, the amount of instructional time, the grouping arrangement (for example, individual instruction versus small-group instruction), or some other aspect of teaching. In this process, the teacher is looking for the type and amount of instruction that will enable the child to make enough progress toward meeting the goal.

**Assessment** is the act of gathering information on a daily basis in order to understand individual student’s learning and needs. **Evaluation** is the
culminating act of interpreting the information gathered for the purpose of making decisions or judgments about student’s learning and needs, often at reporting time. Assessment and evaluation are integral components of the teaching-learning cycle. The main purposes are to guide and improve learning and instruction. Effectively planned assessment and evaluation can promote learning, build confidence and develop students’ understanding of themselves as learners. Assessment and evaluation data assist the teacher in planning and adapting further instruction. As well, teachers can enhance students’ understanding of their own progress by involving them in gathering their own data and by sharing teacher-gathered data with students. Such participation makes it possible for students to identify personal learning goals. The following principles are intended to assist teachers in planning for student assessment and evaluation:

- Assessment and evaluation are essential and integral aspects of the teaching-learning process. They should be planned, continuous activities that are derived from curriculum outcomes and consistent with the instructional and learning strategies.
- A variety of assessment and evaluation techniques should be used and they need to be selected for their appropriateness to both students’ learning styles and to the intended purposes. Students should be given opportunities to demonstrate the extent of their knowledge and abilities in a variety of ways.
- Teachers should communicate assessment and evaluation strategies and plans in advance, informing the students of the outcomes and the assessment procedures relative to the outcomes. Students should have opportunities for input into the evaluation process.

Educational reformers are seeking answers to two fundamental questions: (1) How well are students learning? And (2) how effectively are teachers teaching? Classroom assessment responds directly to concerns about better learning and more effective teaching. Classroom assessment, involves student and teachers in the continuous monitoring of students' learning. It provides faculty with feedback about their effectiveness as teachers, and it gives students a measure of their progress as learners. Most important, because classroom assessments are created, administered and analyzed by teachers themselves on questions of teaching and learning that are important to them, the likelihood that instructors will apply the results of the assessment to their own teaching is greatly enhances. The classroom assessment process assumes that students need to receive feedback early and often, that they need to evaluate the quality of their own learning and that they can help the teacher improve the strength of instruction. Assessment is integral to the
teaching–learning process, facilitating student learning and improving instruction and can take a variety of forms. Classroom assessment is generally divided into three types: assessment for learning, assessment of learning and assessment as learning. Classroom assessment is the process of collecting information from your students about their experience as learners in your class. There are many different ways of collecting information, depending on what you are teaching and what kind of information teacher need.

1.5 ACCOUNTABILITY AND EVALUATION

Students, teachers and administrators have always been held accountable, primarily at a local school level or district level and sometimes at the state level. In the last decade unprecedented accountability testing policy initiatives have increased the pressure on schools to show positive test results. Evaluation is a process used to systematically collect valid and reliable information about a program. This information is then analyzed and interpreted to determine their meaning. Written reports become the visible product resulting from evaluation process.

Like evaluation accountability is also a process. But unlike evaluation, accountability is largely a public information process specifically targeted towards those inside and outside the organization that are in position to influence the decision made about extension program, budgets, staffing, facilities and related matters. And it is important to be aware of internal and external purposes that accountability deserves.

- Internal accountability influences program management decisions.
- External accountability provides concrete evidence of our accomplishments to administrators.

Evaluation can be planned to meet both external accountability needs as well as simultaneously provide information for program improvement or management purposes. Evaluation provides basic facts. Accountability uses these facts to influence the influential. Accountability reporting without solid evaluation backup is hallowing. Just like a doctor should be accountable for his patients, a teacher is responsible for the well-being of the whole child. Because of this, there should be a wide range of evaluative criteria used to give a teacher a formal review at the end of the year. What has the child accomplished in this classroom? Is there a portfolio of his or her work? How far has he or she come in speaking, listening,
reading and writing? What mathematics skills is he or she coming away with? And, perhaps most important of all, does the child leave that class more than ready for the next grade on a social as well as an academic level?

Teacher accountability has to do with knowing best practices (and using them); it entails intimate knowledge of the curriculum, state standards, and having the skills to deliver superior instruction. It also has to do with knowing that what happens in the classroom is not about the teacher but about the student and his or her success, but that success is fluid and should never be tied to one assessment given on one day, rather it should be based on a myriad of things that will gauge performance over an extended period. Being a teacher is truly a calling, and the person who steps into a classroom must take on everything that came before him or her, all that is going on in the present and needs to be aware of all things coming up ahead. Good teachers never stop learning and never stop doing. Evaluations of teachers should include many elements besides state assessment scores, and there should be a direct correlation between the students’ total accomplishments for the year and teacher ratings.

1.6 SELECTION OF APPROPRIATE TOOLS FOR STUDENTS’ EVALUATION

When selecting an assessment tool, it is important to know the type of measures available. Measures generally fall into one of two categories; indirect-measures or direct-measures. Indirect measures examine perceptions relative of an outcome. They may collect information regarding individuals’ perceptions of how well they completed a task or what they feel has been learned. Examples of indirect-measures (adapted from Suskie, 2009) include:
• Reflective essays
• Amount of time spent at extra-curricular activities related to course
• Focus Groups
• Activity Volume
• Job Placement Data

Direct measures examine the actual results of completing a task (e.g., essay, test, exercise and performance) and are evaluated by an expert or authority (e.g., instructor, national testing service). Examples of direct-measures (adapted from Suskie, 2009) include:
• Written work, performances or presentations
• Scores on locally-designed tests or essays (e.g., final examinations)
- Observation of student behavior
- Course/assignments evaluated using a rubric
- Case studies
- Competency-based written and practical tests
- Essays
- Homework assignments
- Quizzes
- Term papers
- Written assignments

While selecting an appropriate assessment tool
- Examine an outcome and its rubric.
- Decide which type of measure is needed (indirect or direct)
- Review assessments already used in your discipline, unit or service area and decide
- If any assessments already exist that addresses the outcome.
- Check what colleagues at other institutions are using

When selecting an assessment tool, please consider the amount of time, resources and support you have to conduct measures. Yes, we want a method of assessment that is both systematic and reliable, but also one that can be completed within a reasonable time frame and with a reasonable amount of effort. An excellent strategy to save both time and energy is opting for embedded assessments. Embedded assessments are measures that are already a part of your course, program or unit. Quite likely, there is no need to create additional surveys, papers, essays, tests, performances or projects to measure learning outcomes if there are already sound assessment tools (measures) in place.

1.7 USE OF EVALUATION FOR INSTRUCTIONAL FEEDBACK

The teacher’s response is called feedback which is the transfer of information from the teacher to the students following an assessment. A simple definition of feedback is confirming the correctness of an answer or action, whether it is right or wrong. This is what we do with most tests by telling students what they got right and what they missed; it is also the extent of feedback many teachers give to a student’s answer to oral questions in the form of “Good”, “That’s right” and so on. Feedback of this nature is part of what students need to improve their learning. Of course feedback is also provided in the form of grades on unit tests and report cards, though normally grades offer very limited feedback.
Evaluation systems face many challenges. In addition to the quality, relevance and timeliness of the evaluation itself, a major challenge lies in conveying the evaluation results to multiple audiences both inside and outside development agencies. Thus feedback and communication of evaluation results are integral parts of the evaluation cycle. Effective feedback contributes to improve development policies, programmes and practices by providing policymakers with the relevant evaluation information for making informed decisions. The need for improved evaluation feedback is widely recognized by development agencies, both on grounds of accountability and learning. Greater accountability is seen as a prerequisite for continued support for development assistance within donor countries, while also being a vital component in creating robust and meaningful partnerships with countries and organisations that are recipients of aid. Better learning is crucial for improving the effectiveness of aid and ensuring that the hard won lessons from experience – both positive and negative – are heeded. Evaluation feedback is an umbrella term describing the various channels and mechanisms by which these crucial messages get through – and make a difference.

Feedback is information about how we are doing in our efforts to reach a goal. Specific, descriptive feedback is necessary for improvement and success. How teachers provide suggestions for improvement is critical in ‘closing the gap’ for students. Teachers who combine strong subject knowledge with effective feedback can offer students rich, focused information about their learning and how to improve it. Students who are clear about their learning can monitor their progress and seek feedback to improve their learning. Professor James Pennebaker from the University of Texas at Austin has been researching the benefits of frequent testing and the feedback it leads to. He explains that in the history of the study of learning, the role of feedback has always been central. He said that when people are trying to learn new skills, they must get some information that tells them whether or not they are doing the right thing. Learning in the classroom is no exception. Both the mastery of content and, more importantly, the mastery of how to think require trial-and-error learning. Feedback is the most effective when it is given at the time of the learning so that students can make improvements as they go. Research literature as well as common experiences, has confirmed that the right kind of feedback is essential for effective teaching and learning. Corrective feedback is needed for learning and motivation and assessment is needed to provide feedback. Feedback is helpful when it has the following characteristics.
Be as Specific as Possible

For example, feedback like "Great job!" doesn't tell the learner what he did right, and likewise, a statement such as "Not quite there yet" doesn't give her any insight into what she did wrong and how she can do better the next time around. Instead, provide learners with information on what exactly they did well, and what may still need improvement. They also note that it can be helpful to tell the learner what he is doing differently than before.

The Sooner the Better

Numerous studies indicate that feedback is the most effective when it is given immediately, rather than a few days, weeks or months down the line. In one study that looked at delayed vs. immediate feedback, the researchers found that participants who were given immediate feedback showed a significantly larger increase in performance than those who had received delayed feedback. Another research project from the University of Minnesota showed that students who received lots of immediate feedback were better able to comprehend the material they had just read. Of course, it's not always possible to provide students with feedback right on the spot, but sooner is definitely better than later.

Address the Learner's Advancement towards a Goal

Timperley and Hattie note that effective feedback is most often oriented around a specific achievement that students are (or should be) working toward. When giving feedback, it should be clear to students how the information they are receiving will help them progress toward their final goal.

Present Feedback Carefully

The way feedback is presented can have an impact on how it is received, which means that sometimes even the most well-meaning feedback can come across the wrong way and reduce a learner's motivation. Psychologist and author Edward Deci has identified three situations in which feedback could be counterproductive:

a. **When learners feel too strictly monitored:** If learners feel that they are being too closely monitored, they might become nervous or self-conscious, and as a result, disengaged from learning.

b. **When learners interpret feedback as an attempt to control them:** Learners may sometimes interpret feedback as an attempt to control them or tell them how they *should* be doing something rather than guidance on how to improve.
c. **When learners feel an uncomfortable sense of competition:** Feedback shared in a group setting could cause learners to feel like they have to compete with their peers. This can be another source of disengagement in learning.

To avoid these situations, Deci suggests fully explaining the purpose of any monitoring and ensuring that learners understand how the feedback is meant to help them compete against their own personal bests rather than each other.

**Involve Learners in the Process**

The importance of involving learners in the process of collecting and analyzing performance based data cannot be understated. Students **must** be given access to information about their performance. At the broadest level, students need to know if they actually have mastered the material or not. Giving them information about the ways they are studying, reading, searching for information or answering questions can be invaluable. When students have access to this information, they develop an awareness of their learning and are more easily able to recognize mistakes and eventually develop strategies for tackling weak points themselves.

**Feedback should be Educative in Nature**

Providing feedback means giving students an explanation of what they are doing correctly and incorrectly. However, the focus of the feedback should be based essentially on what the students is doing right. It is most productive to a student’s learning when they are provided with an explanation and example as to what is accurate and inaccurate about their work.

**Be Sensitive to the Individual Needs of the Student**

It is vital that we take into consideration each student individually when giving feedback. Our classrooms are full of diverse learners. Some students need to be nudged to achieve at a higher level and other needs to be handled very gently so as not to discourage learning and damage self-esteem. A balance between not wanting to hurt a student’s feelings and providing proper encouragement is essential.

**Give Feedback to Keep Students “On Target” for Achievement**

Regular ‘check-ins’ with students lets them know where they stand in the classroom and with you.
Host a one-on-one Conference

Providing a one-on-one meeting with a student is one of the most effective means of providing feedback. The student will look forward to having the attention and allows the opportunity to ask necessary questions. A one-on-one conference should be generally optimistic, as this will encourage the student to look forward to the next meeting. As with all aspects of teaching, this strategy requires good time management. Try meeting with a student, while the other students are working independently. Time the meetings so that they last no longer than 10 minutes.

Feedback can be Given Verbally, Non-verbally or in Written Form.

Feedback can be given verbally, non-verbally or in written form. Be sure to keep your frowns in check. It is imperative that we examine our non-verbal cues. Facial expressions and gestures are also means of delivering feedback.

Concentrate on the one Ability

It makes a far greater impact on the student when only one skill is critiqued versus the entire paper being the focus of everything that is wrong. For example, when I taught Writer’s Workshop at the elementary level, I would let students know for that day I was going to be checking on the indentation of paragraphs within their writing. When I made conference with a student that was my focus instead of all the other aspects of their writing. The next day would feature a new focus.

Alternate Due Dates for your Students/classes for Effective Feedback

Utilize this strategy when grading papers or tests. This strategy allows you the necessary time to provide quality, written feedback. This can also include using a rotation chart for students to conference with at a deeper more meaningful level. Students will also know when it is their turn to meet with you and are more likely to bring questions of their own to the conference.

Give Specific and Descriptive Feedback

It is important to be as specific and descriptive as possible when giving feedback. If the feedback is vague or general, it will not be helpful to the students; it will communicate only a sense of goodness or badness of the performance. A descriptive statement specifies in exact terms the nature of the performance of the performance.
Focus Feedback on Key Errors

It is not practical to provide detailed and specific feedback to every student on homework and other assignments. It is best to determine what the most significant error is or what changes will be most helpful to the students.

1.8 SELF ASSESSMENT QUESTIONS

Objective types
1. The systematic process of collecting and analyzing data in order to make decision is called:
   a. Evaluation
   b. Measurement
   c. Assessment
   d. Testing
2. The process of quantifying the degree to which someone or something possesses a given trait is called:
   a. Evaluation
   b. Measurement
   c. Assessment
   d. Testing
3. Identify five problems due to lack of evaluation in schools.
   1. ____________________________________________
   2. ____________________________________________
   3. ____________________________________________
   4. ____________________________________________
4. Which type of problems can emerge if students are not evaluated properly in schools?
5. How does evaluation play role in students’ life?
6. Why feedback is linked to evaluation? Give arguments to clarify your point of view.
7. Selection of an appropriate tool for students’ evaluation is the major decision for accurate evaluation. Discuss.
8. What is the difference between evaluation and accountability? Give an example to support your point of view.
Subjective types:

Q.1 What does evaluation mean to the school? Why is it conducted?
Q.2 How does the school plan and implement student evaluation?
Q.3 Discuss the function of evaluation. Why is evaluation complicated?
Q.4 Explain the need of evaluation. How does evaluation affect students’ life?
Q.5 Analyze the methods to evaluate the students?
Q.6 Accountability and evaluation are similar in nature? Why? Why not?
Q.7 Why should the teachers be careful in their evaluations?
BIBLIOGRAPHY


NEED, SCOPE AND KINDS OF TESTS

Written By: Dr. Munnaza Ambreen
Reviewed By: Dr. Muhammad Tanveer Afzal
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>25</td>
</tr>
<tr>
<td>Objectives</td>
<td>25</td>
</tr>
<tr>
<td>2.1 Purpose, Need and Scope of Test</td>
<td>26</td>
</tr>
<tr>
<td>2.2 Purpose and Concept of Testing</td>
<td>27</td>
</tr>
<tr>
<td>2.3 Test Types and Their Use in Educational Measurement</td>
<td>29</td>
</tr>
<tr>
<td>2.4 Significance of Objectives and Subjective Type Tests</td>
<td>34</td>
</tr>
<tr>
<td>2.5 Example of Renowned Tests and Their Use</td>
<td>35</td>
</tr>
<tr>
<td>2.6 Self-Assessment Questions</td>
<td>37</td>
</tr>
<tr>
<td>Bibliography</td>
<td>38</td>
</tr>
</tbody>
</table>
INTRODUCTION

Classrooms are places of diverse learners. Students of different learning capacities and individual differences come in class. Classroom tests play very vital role to judge the effectiveness of teaching learning process and student achievement. Classroom assessment strategies and techniques also help the teachers and students to point out the deficiencies and help to improve teaching and learning. Classroom tests create interest in studies and develop sense of competition among students. Classrooms are black boxes and to know what is going in class whether objectives are achieved or not, tests play central role to know the effects of classroom activities.

Tests play major role to ensure and enhance the effectiveness of teaching learning process. Teachers need to get information about students’ performance. In order to increase the motivation and enhance students’ performance; it is needed formal and systematic assessment, which is possible with the help of tests. On the bases of information, teachers take decision or necessary measures for student betterment. To identify the weakness and strengths of students’ attainment, to judge the effectiveness of instructional techniques and to take better decision about students welfare, it is only possible when valid information is available and it can be achieve through tests.

OBJECTIVES

After studying this unit, the students will have ability to demonstrate.

- understand and describe the purpose, need and scope of tests.
- elaborate the concept of testing.
- understands the types of tests and their role in educational measurement.
- identify the significance of objective and subjective type tests.
- describe the role of renowned tests and their use of educational field.
2.1 PURPOSE, NEED AND SCOPE OF TEST

Need of Test
The trend of testing is rapidly increasing in educational institutions now days. It can very valuable for students for motivating and creating interest in studies. The result of tests compels the students for better performance. Sometimes sub-testing, judges the progress of students. You know that strength of students is increasing in school and colleges day-by-day, so it is introducing organize system of testing to evaluate the performance of students. Tests are very useful device for making better admission in educational institutions. Tests affect the performance and task of teachers. Teachers have complete awareness of using these devices because it is very helpful for successful educational program so that when the teachers interact with stakeholders like students, teachers and community confidently.

It is fact that teaching learning process is incomplete if teacher have not use properly this effective device because he cannot measure the performance of students without this device. It is most important for teachers to know how their class going on and which are the aspects of educative process, improvement is needed. Ultimately you have better understanding of importance of tests. It is beneficial for us to know about test.

You people observe that teachers use oral and written tests to taking information about classroom activities and student achievement for years. Teachers use tests to assess the effectiveness of teaching learning process and take remedial steps for improving his teaching. It is very effective tool in the hand of teacher. A teacher made tests is used to collect valid information about students’ performance and this information about students’ weakness and strength are being reported to parents.

Education for all is almost slogan of all countries. Any institution cannot refuse for student’s admissions. Then student’s placement is big challenge for new students without tests .So selection tests are being conducted for all types of students for their proper place. These tests are very useful for student placement because results of selection tests indicate the student proper level and place in institution.

With the help of tests, students can be classified on the bases of their ability. Students can be divided into groups. Teacher may fully aware of students’ level and he can make his teaching on the mental level of students and make his teaching more interesting and easier for students. It can provide fruitful result for teachers concerning students’ performance.
A teacher can assess his teaching through the results of tests. Instructional objectives can be achieved only when teachers involve the students in teaching learning process and interact with them in class effectively. Use of test in class is very important because teacher can assign grade to students and student can be motivated in this way and sense of competition can be promoted among students.

2.2 PURPOSE AND SCOPE OF TEST

Now we discuss scope and purpose precisely. Classroom is place of learning for students. School provide healthy environment for holistic development of students. Classroom assessment of students is inevitable. It is most important part of educational programs how students are learning. It helps in judging the effectiveness of instructional program. Assessment not only provides information for teachers to refine their teaching but also it is much supportive for better learning. There are different types of assessment techniques are being used by the teacher to evaluate the performance of students such as observation, tests, questionnaires and interview. Teacher design the tests to get numerical data about students performance in subject matter or measure student behavior in class.

Selection Decisions

Test is used to enter any institution because it is appropriate tool for selection of students among thousands of students on the basis of test result. It is impossible to get admission all students in institution. Different tests and techniques are used to select the most suitable candidates. Different tests such as achievement tests, intelligence tests and aptitude tests are being used to select suitable candidates. On the basis of test results, students are being selected or rejected for different programs of school, college and university.

Classification of Students

A classroom teacher has prime importance in educative process. As assessor, he has to take important decision concerning about students’ placement and their classification on the basis of test results. Students can be classified on the basis of their ability high, average and low ability. Test score helps the teachers in effective way that he can classify into different groups according to mental, emotional and learning level of students.

Grading Decision

With the help of test, teacher take grading decision on the basis of test result. He can assign grade to student on objective information of student learning and academic achievement. Teacher made test are used to grading decision about
student and they can categorize the students in A, B, and C, grade. It is responsibility of teacher to assign grade to students on the basis of objective data.

**Monitor Students Performance**

To know the effectiveness of teaching, it can be assessed, the performance of students is satisfactory or not. Appropriate use of test provides strong information of students’ progress. It can be judged through classroom tests that students are going towards for achieving objectives. Teacher can get information about achieving objectives after that he can bring positive change in his teaching method. If students are achieving instructional objective then there is no need to change instructional procedure. The results of classroom tests can provide feedback to teacher and students.

**Diagnosis and Remedial Decisions**

Tests and measurement data play key role to identify the strength and weakness of students. The teacher takes different positive steps during the phases of teaching for effective learning. In spite of these encouraging steps to make better teaching if students have any problem in learning, comprehension of concepts is creating problem, teacher will feel this problem and try to find out weakness through improving his teaching according to satisfaction level of students. Because good teaching is diagnosed learning difficulties of students and take remedial steps for improving teaching learning process in effective way.

**Guidance and Counseling Decision**

The role of guidance and counseling for student welfare has key importance. The test data also help to take better steps regarding guiding and counseling of students. The result of different test battery like different aptitude test also help the teacher for guiding on proper direction and it helps in choosing the right path according to their will.

**Concept of Testing**

As you know that test is formal and systematic tool to collect information of student behavior. In education, a test consists of a question or series of questions or exercises or other devices for measuring the mental ability, capacity, skill, knowledge, achievement, performance, aptitude, attitude, interest, social and emotional adjustment of personality of individual and group.
A test is an effective device which is frequently used in schools to measure the behavior of person for particular purpose. It is a tool to check the learning task of students. It is an effective device for measuring performance of students in numerical form. You must keep this thing in your mind that it is not possible for teachers that he includes all content in test items but he tries to include all important and maximum content for preparing test. So that he may cover maximum content. It is reality that you cannot cover all topics and concept for assessment of two hour. In beginning, paper pencil tests were being used but with the passage of time, different other types of tests such as subjective types and objective type tests are being used in educational institutions to evaluate the students’ performance. Testing and measurement denote the same concept of obtaining a numerical description of the degree to which an individual possess a particular characteristic (Grunlnd, 2006).

Test is limited in scope because it consists of series of questions. Tests are quantitative tool that are classified in measurement. In measurement, the data is taken in numerical form after items response towards the students and interpretation of numerical data is assessment. When the tests are administrated to evaluate the students performance at the end, it is summative evaluation while to assess the learning during instruction is formative. Tests, measurement and evaluation techniques are used for purpose of evaluating students but tests measure the effectiveness of classroom practices

2.3 TEST TYPES AND THEIR USE IN EDUCATIONAL MEASUREMENT

To evaluate student’s performance is key element of assessment program. Assessment is an integral part of educative process. In the light of test results several decisions have been taken by educational manager, classroom supervisors, educational expert and classroom facilitator to enhance the students performance, to improve curricula, to refine pedagogical skills for healthy growth of coming generation. For this purpose classroom tests are very effective tool. To take better and fruitful decision is only possible when we have information about teaching learning practices and this information is only achieved through tests.

Today, there are several types of tests that it is hardly possible to classify them into different categories. For example, norm and criterion referenced tests, power and speed tests, aptitude tests personality tests, intelligence tests and tests for individual and group. These tests are being used according to need and situation. There is great significance of each type test in the field of educational assessment. Diagnostic tests are used to find the weakness and strength.
Standardized and teacher made tests are very popular in field of educational assessment now-a-days. Commercially developed and produced tests which have been normed against standard population are known as standardized tests. These tests are highly reliable and valid and use to measure students performance on large scale. It is time consuming and expensive. The tests used by the teacher in daily classroom practice to assess student achievement level in concerning of achieving objectives are called teacher made tests. These test provide feedback that they can assess students’ performance, their interest in studies and result of these tests take attention of teachers to make his teaching interesting and effective. We discuss here subjective type and objective type tests. These tests can be further subdivision as follows.

1. **Essay Type Tests**
   a. Extended Response Items
   b. Restricted Response Items

2. **Objective Type Tests**
   a. Supply Type
      i. Completion Items
      ii. Short Answer
   b. Selection Type
      i. Matching Items
      ii. True False Items
      iii. Multiple Choice Questions

**Essay Type Tests**

In essay type tests, respondent response freely but in lengthy way. Through essay type tests, it is very easy to measure high level thinking skills such as applying, organizing and evaluating. It is not appropriate to measure knowledge level outcome. Students can integrate and organize their ideas in forms of paragraphs of many pages in detail and teacher can evaluate the quality and accuracy of their response.

Essay questions provide clear picture of complex learning outcomes. Creativity and expression of ideas in written form developed effectively through this type tests. It is easy to prepare and coping is not possible. It can measure complex learning outcomes. Essay type tests are being used at the end of semester or take home tests in flexible page limits maximum range of thirty pages. There are two main types of essay type tests restricted response questions and extended response questions.
I. Restricted Response Questions

Restricted response questions are always restricted in terms of content and response. It measures more specific learning outcomes and more objective than extended response questions. Example

i. Describe qualities of good teacher.
ii. Explain the causes of ill literacy in Pakistan.

Advantages

It is more reliable and objective than extended response and covers wide area of content. Scoring is easy and quick as compared to extended response.

II. Extended Response Questions.

Extended response questions allow the students to response freely in extended form in several paragraphs and answer the question in best way and organize their ideas effectively. Example

i. What measures can be taken to enhance quality of education in Pakistan.
ii. Evaluate the agricultural development of Pakistan and correlate it to the economic development of Pakistan.

Advantages

These type questions provide complete freedom to response. Writing power of students develops and arguments presentation ability is created among students. They demonstrate the ability of selecting and organizing ideas, it can be constructed very quickly but scoring is subjective.

Objective Type Tests

Objective type tests are very popular now-a-day. They cover lot of content area. There are many forms of objective type tests. It develops all mental capacities of students because students have to select one correct answer in given options. The scoring of items is objective and biases of examiner have no affect on marking. The construction of items is needed skilled person. There are some rules and techniques to prepare the items. There are two types of objective type tests.

1. Supply Type
2. Selection Type

1. Supply Type

A. Completion Items

Completion items are easy to write. It is in form of sentence in which one or more words have been left out with blanks. The student fills blanks with
correct word. It is presented in incomplete statement. It is most effective for assessing knowledge and comprehension level learning outcomes. For example
i. The capital city of Pakistan is…………..

**Advantages**
Construction of question is easy. Guessing is eliminated due to recalling. Greater portion of content can be covered and take less time to complete than multiple choice items

**Disadvantages**
Completion questions usually encourage relatively low level of response complexity. And it is comparatively difficult to score.

B. Short answers
The short answer items require pupils to supply appropriate word, number or symbol to direct question or incomplete statement. It is very effective to measure knowledge and comprehension level outcomes but can be written for higher level outcomes. Example
i. The name of first Governor General of Pakistan is.

**Advantages**
They are relatively easy to construct for teachers. It reduces the possibility of guessing. It covers greater amount of content and asses lower level learning outcomes.

**Disadvantages**
Scoring is tedious and time consuming. Short answer items provide little opportunity for students to synthesis, apply and evaluate.

2. **Selection Type**

A. Matching Items
Corey (1988) defines matching items; “Matching test items are another popular selected response format. These items require students to match information in two columns. Items in left hand column are called premises and those in right hand column are called responses. Students are required to locate correct response for each premise” It consists of two parallel columns. Word, symbol, number in one column is matched to a word, sentence or phrase of other column.
Advantages
Matching items are usually simple to construct and score. They are ideally suited to measure associations between facts. It reduces affects of guessing.

Disadvantages
Matching items emphasizes on memorization. It can assess only factual information. It provides clues.

B. True False Items
Nitko (1983) defines, “A true false item consists of statement or proposition which examinee must Judge and mark as either true or false” There are two answers against each statement right and wrong whatever form is given to it, if one answer is true and other is false, it can be called true/false test. Example
I. Pakistan is a developed country. T/F
II. Lahore is in Punjab. T/F

Advantages
It takes less time of construct. It covers wide sampling of content. Scoring is mechanical so high degree of objectivity is possible. It measures simple learning outcomes.

Disadvantages
It allows high degree of guessing (50 percent) as there are only two choices. Copying is possible. It emphasizes rote memorization of knowledge and cannot measure complex learning outcomes.

c. Multiple Choice Questions
Gilbert Sax (1989) defines, “Multiple choice items consists of two parts; a stem and number of options or alternatives. Stem is question or statement that is answered or completed by one of the alternative. All incorrect or less appropriate alternative are called distracters or foils and students task is to select the correct of best alternative from all options”

Advantages
Multiple choice questions have considerable versatility in measuring objectives from knowledge to evaluation level. Since writing is minimized. A substantial amount of course material can be covered in short time. The scoring is highly objective. It reduces chance of guessing. Bluffing is not possible.

Disadvantages
These items are time consuming to write. They can check superficial and limited knowledge. They cannot measure attitudes and motor skills. Intelligent examinees
suffer if they choose incorrect answer due to ambiguities of words while poor examinees choose correct answer due to their limited knowledge.

2.4 SIGNIFICANCE OF OBJECTIVE AND SUBJECTIVE TYPE TESTS

Classroom tests have greater importance in an educational system. These tests provide valid and reliable data about student’s performance. They indicate the weakness and strengths of teaching and the information of tests are very useful for teachers as well as students. The success of students is considered the teachers success. They provide feedback to teachers and necessary measures are being taken in the light of test results. Teachers improve their teaching and students also take much interest in studies because results of tests are like mirror for students.

Subjective and objective type tests both have their own significance in field of educational assessment. Objective type tests assess learning outcome of knowledge level of cognitive domain because they measure the ability of remembering some terminology, facts and principles. They hardly test other levels of cognitive domain like comprehension, application and evaluation. On other hand, subjective type tests assess complex learning outcomes such as application, analysis, synthesis and evaluation. In spite of their drawbacks, they are necessary for healthy educational system due to measuring reasoning and intellectual abilities of students. As you know both type of tests has their own importance in present era. So, we discuss the importance of both types precisely.

Significance of Subjective Type

The subjective items demands fully developed response from students. The student’s response the subjective question in several paragraphs in appropriate and effective way. The quality and accuracy of response assess the subject specialist who has good command on subject. Essay items test abilities of logical thinking, critical reasoning and effective presentation of ideas. These tests allow students to show their writing expression abilities, arguments and originality of ideas in free response. They develop study habits among students. The students prepare outline, summaries and organize the topic in impressive way. These tests are being used to evaluate all subjects. They develop among students analytical viewpoint problem solving skills and enhance divergent thinking.

Students compare, contrast and evaluate different topics in effective way and present valid arguments to prove their answers with facts logically. Through essay type tests, students opinion and analysis of different issues and topics can be seen
due to written expression in form of sentences. They measure complex learning outcomes because they allow the students to response freely and put ideas in favor of their response.

Subjective type tests take less time to construct and very easy to prepare and students respond essay type questions in comprehensive way. Sequence and organization of ideas defending power of arguments in favor of response develop among student through these types of tests and they develop mental faculties of students.

**Significance of Objective Type**

An effective assessment is essential part of education system. In present era, the significance of objective type test in daily assessment is prime element of educational institutions. Assessment provides valid information about student’s performance and necessary measures are being taken after results of the tests. The success of good teaching, overall quality of good teaching and reputation of institution can be judged possible when you have reliable data of student’s performance and objective type tests provide objective and authentic information of student’s achievement. This information is correct because biases of examiner have no space in scoring. Even untrained person can assess accurately with the help of already prepared key. So objective type tests provide clear and unambiguous scoring criteria.

These tests cover wide area of content. Students cannot respond if they prepare selective study. They have to study over all content and they are able to select the correct answer within given options. The decision power of students enhance and study habits also develop among students. Pupil likes this type of test because there is no chance for teachers to show favoritism and students take interest in answering these tests. These tests discourage cramming and encourage thinking.

These tests can be used to assess students performance in all subjects of school and judgment of these tests are easy and quick providing quick feedback to students. High reliability, objectivity and adequate content sampling is quality of objective type tests. Written expression do have effect on scoring. They are most suitable for measuring lower level outcomes like knowledge of facts and dates but some other types also measures other aspects of complex learning outcomes.

**2.5  EXAMPLE OF RENOWNED TESTS AND THEIR USE**

A test is an assessment intended to measure test taker knowledge, skill, aptitude and physical fitness. These are being used to aware of student’s performance. Some countries such as United Kingdom and France require all their secondary
school students to take standardized test on individual subjects such as General Certificate of Secondary Education (GCSE). As you know that there are many types of tests but we discuss some popular test and their use in assessment.

**SAT**

It is a standardized test widely used for college admission in United States. It was first used in 1926 but its name and scoring criteria have changed several times. It is called Scholastic aptitude test and then its name changed and it is called Scholastic Assessment Test. The present SAT has been introduced in 2016. The SAT is published by the College Board, a private and non-profitable organization in United States. It is prepared and administered on behalf of College Board by Educational Testing Service. The test is prepared to know the students placement and readiness for college.

SAT consists of three major sections; critical reasoning, mathematics and writing. It measures literacy and writing skills to achieve success academically in colleges. SAT assess how test takers analyze and solve the problems skills that they learn in school that they need in college. Test is taken in limited time to produce arrange of score. It is compared to the performance of high school with college success. There is prominent differences in funding, curricula and grading among united states secondary school due to local control, private, distance and home schooled students. SAT scores provide admission authorities to put local data in a national perspective. SAT is also used to assess students’ placement and admission in colleges and assess students performance in reasoning, writing and mathematics.

**Graduate Record Examinations (GRE)**

GRE is the most popular test that is used to take admission in most graduate schools in United States. It was prepared and administered by the Educational Testing Service in 1949. It measures verbal reasoning, quantitative reasoning, analytical writing and critical thinking skills that have been acquired over long period of time and that are not entirely based on specific field of study outside of the GRE itself. The importance of GRE is vital and it is used to selection of students in universities.

Many graduate schools in USA is adopted this exam for admission process and it is used in most of countries of world for admission. In Pakistan it is also used for admission in universities. It is standardized test to measure general abilities of students as well as specific ability of subject’s specialization. Unlike other admission tests, the worth and value of GRE is prominent from school to school, department to department and from program to program. This test also use in all
types of discipline such as agriculture, economics, business, science, engineering and liberal arts for admission purposes. And it is integral part of admission process; many countries are using this test in admission of their higher level institutions because they measure the student’s academic ability in their field area and the most popular test to get admission in universities.

IOWA Tests of Basic Skills

This is standardized test known as IOWA test. It was designed and developed in 1935 and it is helpful in improving student instruction. For many years, it has been being used to asses’ students’ performance in United States. School participate in these tests annually and improve their program. It has been revised in 2011-2012 for improvement in school.

The IOWA test of basic skills is given as standardized to children ranging from kindergarten to 8th grade. It is an achievement test and assess children knowledge what they have learned in school. Students are tested on vocabulary, word analysis, reading comprehension, language and social skills. A series of test in school are being used to gain information about classes, student performance to supplement teacher observation regarding student’s abilities and provide base of student evaluation. Student’s skills are being developed and the result of the test provides help to improve curricula and lesson planning.

It is very effective test to asses’ student’s skills, knowledge and learning level in their school. And this test betterly provide information of student’s performance individually. It motivates and develops habit of study among students. It evaluates student’s performance in key academic areas and decision has been taken about instructional planning and curricula.

2.6 SELF ASSESSMENT QUESTIONS

1. Describe the role of these tests in our educational system.
2. Describe how classroom test is useful for students.
3. Why essay type is necessary?
4. Describe significance of objective type tests.
5. Describe the significance of subjective type test in present sanerio.
6. Highlight the role of objective type tests in enhancing student’s performance.
7. Compare and evaluate essay type and objective type tests by giving comparable merits and demerits of both tests.
BIBLIOGRAPHY


CLASSROOM TESTING AND HIGH STAKE TESTING

Written By: Inam Jahingir
Reviewed By: Dr. Naveed Sultana
## CONTENTS

- Introduction ............................................................................................................. 41
- Objectives .................................................................................................................. 41

### 3.1 Concept of Classroom Testing and Its Techniques ................................................... 42
  - 3.1.1 Classroom Assessment Concepts and Applications ........................................ 42

### 3.2 High Stake Testing: Its Nature, Need and Scope .................................................... 44
  - 3.2.1 What is Need of High Stake Testing ............................................................. 45
  - 3.2.2 What is the Nature of High Stake Testing ..................................................... 46

### 3.3 Teacher Made Tests ............................................................................................. 46

### 3.4 Standardized Tests ............................................................................................. 48

### 3.5 Benefits and Limitations of Tests ......................................................................... 49
  - 3.5.1 Advantages and Limitations of Teacher Made Tests ...................................... 50
  - 3.5.2 Advantage and Disadvantage of High Stake Testing ................................. 51

### 3.6 Concept and Use of Taxonomies in Test Development ........................................... 52
  - 3.6.1 Using Bloom’s Taxonomy in Test Development .......................................... 53
  - 3.6.2 Using SOLO Taxonomy in Test Development .............................................. 56

### 3.7 Procedure or Steps for a Standardized Test Development Process ....................... 58

### 3.8 Examples of Standardized Tests with Characteristics .......................................... 59
  - 3.8.1 Reliability of Tests ..................................................................................... 61
  - 3.8.2 Validity of Tests ....................................................................................... 63
  - 3.8.3 Usability of Tests ..................................................................................... 65

### 3.9 Current Trends of High Stake Testing .................................................................... 65

### 3.10 Summary of the Unit ......................................................................................... 66

### 3.11 Self Assessment Questions ................................................................................. 67

- Bibliography ............................................................................................................... 68
INTRODUCTION

Assessing student learning is something that every teacher has to do quite frequently. Written tests, book reports, research papers, homework exercises, oral presentations, question-and-answer, science projects and artwork of various sorts are ways in which teachers’ measure student learning. Written tests are counted for about 45 percent of a typical student's course grade (Green & Stager, 1986/1987). Stiggins (1994) holds that a typical teacher spends between one-third and one-half of her class time engaged in one or another type of measurement activity. Despite the amount of time spent by the teachers in assessment, most of them dislike and a few do it well. Teachers with more training in assessment use more appropriate assessment practices than do teachers with less training (Green & Stager, 1986/1987). Classroom Assessment are teacher made tests which are aimed to measure academic achievement, identify learning problems or inform instructional adjustments, among other purposes that is why they are also called “teacher made tests or low stake tests”

A high-stakes test is different not in form (how the test is designed) but in its function (how the results are used) also. For example, if test results are used to determine an important outcome, such as whether a student receives a high school diploma, the test would be considered a high-stakes test regardless of whether it’s a multiple-choice exam, an oral exam or an essay exam. Low-stakes tests generally carry no significant or public consequences—the results typically matter far more to an individual teacher or student than to anyone else—and scores are not used to burnish or tarnish the reputation of teachers or schools.

A basic objective of this chapter is to help the prospective teacher understand how assessment can be used to reinforce and enhance rather than work against, our role as teacher and enhance his understanding about the classrooms assessment and high stake testing.

OBJECTIVES

After studying this unit, the students will be able to:
1. understand the concept of class room testing and its techniques
2. understand the need and scope of high stake testing
3. differentiate between teacher made tests/classroom tests/low stake tests and standardized/high stake tests
4. enumerate advantages and limitations of the low stake and high stake tests
5. prepare tests using Bloom’s Taxonomy and SOLO Taxonomy
6. elaborate the procedure for test development
7. provide examples of standardized tests with characteristics with examples.
8. enlist few trends in high stake testing
3.1 CONCEPT OF CLASSROOM TESTING AND ITS TECHNIQUES

Classroom assessment is the process, usually conducted by teachers, of designing, collecting, interpreting and applying information about student learning and attainment to make educational decisions. There are four interrelated steps to the classroom assessment process. The first step is to define the purposes for the information. During this period, the teacher considers how the information will be used and how the assessment fits in the students' educational program. The teacher must consider if the primary purpose of the assessment is diagnostic, formative or summative. Gathering information to detect student learning impediments, difficulties, or prerequisite skills are examples of diagnostic assessment. Information collected on a frequent basis to provide student feedback and guide either student learning or instruction are formative purposes for assessment and collecting information to gauge student attainment at some point in time, such as at the end of the school year or grading period, is summative assessment. The next step in the assessment process is to measure student learning or attainment. Measurement involves using tests, surveys, observation or interviews to produce either numeric or verbal descriptions of the degree to which a student has achieved academic goals. The third step is to evaluate the measurement data, which entails making judgments about the information. During this stage, the teacher interprets the measurement data to determine if students have certain strengths or limitations or whether the student has sufficiently attained the learning goals. In the last stage, the teacher applies the interpretations to fulfill the aims of assessment that were defined in first stage. The teacher uses the data to guide instruction, render grades, or help students with any particular learning deficiencies or barriers.

3.1.1 Classroom Assessment: Concepts and Applications

Hundreds of books and articles on classroom assessment have been written, but most, if not all, ascribe to an assessment framework articulated in the 1930s and 1940s by Ralph Tyler (1949), who believed that assessment was an integral component of curriculum and instruction planning. Tyler developed a multistep model of curricular and instructional design that began with consideration of what the educator expected the student to be able to know and do after teaching had occurred. He termed these end results of education, “instructional objectives,” which he stated should be crafted by considering both the mental skill, such as “applies” or “creates,” and the subject matter content the student will develop. Good planning, according to Tyler, involved developing a table that specifies the body of objectives students will develop during the course of a school year,
semester or lesson. After the instructional objectives are formulated, educational experiences can be developed that encompass the teaching materials and instructional opportunities that will be provided to students. Also during this planning stage, teachers must consider how they will determine if students have attained the instructional objectives. Indeed, good objectives are those that clearly define the type of activity the students will accomplish to indicate the degree to which the students have attained the objective. After students experience the learning opportunities provided by the teacher and after assessment has occurred, the teacher's task is to examine the assessment results and decide whether students have sufficiently reached the objectives. If they have not, the teacher can revise the educational experiences until attainment has occurred. Thus, Tyler's model of testing emphasized the formative role of classroom assessment. Tyler did not organize the mental skills that make up objectives in any meaningful way. Benjamin Bloom, who earlier was a graduate student of Tyler at the University of Chicago, orchestrated a committee during the 1950s to develop a Taxonomy of Educational Objectives (Bloom et al., 1956). The committee organized mental, or intellectual, skills in a hierarchical fashion from the most basic levels, knowledge and comprehension, to the most advanced levels, applications, analysis, synthesis and evaluation. The Taxonomy has been widely used to organize the types of objectives students of all ages are expected to attain in schools worldwide.

Selected- and Constructed-response Formats. Teachers have an array of item formats upon which to measure student attainment of objectives (see Linn & Miller, 2005; Oosterhof, 2003).

Assessment items can be classified into two categories: selected- and constructed-response formats. It is the student's duty in selected response items to choose one or a few correct options among multiple alternatives. Examples of selected response item formats include multiple-choice, ranking of options, interpretive exercises, matching, true-false, alternate-choice, embedded alternate-choice, sequential true-false and checklists. In constructed-response items, students must supply an answer to a question prompt. Short answer and essay items are common constructed-response items.

Essay items can require students to write either extended or restricted responses. Responses can be restricted by limiting the amount of space available to supply the answer, dictating the number of acceptable answers (“state three reasons …”), or by qualifying in the prompt the expected response length (“briefly describe …”). Restricted-response essays are useful for measuring student attainment of factual knowledge and basic comprehension. Extended-response essays are more appropriate if the goal is to measure students' skills at analyzing, synthesizing, constructing, or evaluating information because they offer students greater
latitude in how to organize and present their thoughts. Performance assessments are another type of constructed-response item. With this format, students are expected to perform an activity or set of activities. They can be asked to perform a process, such as delivering a public speech, or produce a product, such as a science notebook or work of art. Many performance assessments, but not all, attempt to represent real-life contexts or applications and are therefore considered authentic assessments. Because students perform activities during these assessment tasks, performance assessments can be integrated well with regular instructional activities. Constructed-response items must be scored by a judge, using either a norm- or criterion-referenced scoring procedure. In norm referencing, the teacher compares the quality of a student's response to a reference group, which might include the other students currently in the class or to prior students the teacher has taught. The teacher then assigns a score to the student's response based on how the response ranks or where it falls in the distribution of responses in the reference group. Criterion-reference scoring involves basing a student's score on the degree to which the student has demonstrated the attainment of specified knowledge or skills. Academic standards stipulate what students should know and be able to do and performance standards specify the degree to which they have mastered the academic expectations. The criteria or expectations often are defined in a scoring rubric, which provide descriptions of responses on a scale. Teachers can use either holistic or analytic scoring rubrics to render criterion-referenced scores. An analytic rubric allows the teacher to score the constructed response on separate and multiple dimensions, such as organization, accuracy, and voice. For holistic scoring, the teacher produces one overall score. A holistic rubric could be based on multiple dimensions, but the teacher considers all of the dimensions simultaneously to yield the score. Analytic rubrics are more useful if the goal is to provide more extensive and deeper feedback to the student, because the student gets separate scores on multiple dimensions. Holistic scoring takes less time, typically, because only one score per response is made. It works, however, only when there is a high relationship among the dimensions for the responses. For example, if students who are high on organization also tend to be high on accuracy and voice, then holistic scoring can work effectively.

3.2 HIGH STAKE TESTING: ITS NATURE, NEED AND SCOPE

High-stakes testing has consequences attached to the results. For example, high-stakes tests can be used to determine students’ promotion from grade to grade or graduation from high school (Resnick, 2004; Cizek, 2001). While Low-stakes testing has no consequences outside the school, although the results may have classroom consequences such as contributing to students’ grades. Formative assessment is a good example of low-stakes testing.
A high-stakes test is a test, which could result in important consequences for the test taker. Passing that test will bring some important benefits, such as a secondary school certificate, a scholarship, driving license or a license to practice a profession. Failing has important disadvantages, such as being forced to take remedial classes until the test can be passed, not being allowed to drive a car or not being able to find employment.

The use and misuse of high-stakes tests are a controversial topic in public education, in advanced countries and even in Pakistan as they are used not only to assess students but in attempts to increase teacher accountability also.

Precisely we can say that a high-stakes test is a test that:

- is a single, defined assessment,
- has a clear line drawn between those who pass and those who fail, and
- has direct consequences for passing or failing (something "at stake").

For example a pre-medical or pre engineering student in Pakistan appearing in ECAT (Engineering College Admission Test) or MCAT(Medical College Admission Test), a masters students seeking admission in MPhil appearing in GAT(General), SAT, GRE, IELTS, TOEFL etc.

Activity 3.1: Briefly discuss in your group Types of High Stake Testing Used in Pakistan; Identify and enlist them?

3.2.1 What is Need of High Stake Testing?

A high-stakes system may be intended to benefit people other than the test-taker. For professional certification and licensure examinations, the purpose of the test is to protect the general public from incompetent practitioners. The individual stakes of the medical student and the medical school are, hopefully, balanced against the social stakes of possibly allowing an incompetent doctor to practice medicine.

A test may be "high-stakes" based on consequences for others beyond the individual test-taker. For example, an individual medical student who fails a licensing exam will not be able to practice his or her profession. However, if enough students at the same school fail the exam, then the school's reputation and accreditation may be in jeopardy. Similarly, testing under the FPSC has no direct negative consequences for failing students, but potentially serious consequences for provinces who do not strive for educational excellence so they lose their representation in federation. Similarly the stakes like loss of
accreditation, funding, teacher pay, teacher employment, or changes to the school's management, are high for the school, but low for the individual test-takers.

3.2.2 What is Nature of the High Stake Testing?

Any form of assessment can be used as a high-stakes test. Many times, an inexpensive multiple-choice test is chosen for convenience. A high-stakes assessment may also involve answering open-ended questions or a practical, hands-on section. For example, a typical high-stakes licensing exam for a medical technician determines whether the nurse can insert an I.V. line by watching the nurse actually do this task. These assessments are called authentic assessments or performance tests.

Some high-stakes tests may be standardized tests (in which all examinees take the same test under reasonably equal conditions), with the expectation that standardization affords all examinees a fair and equal opportunity to pass. Some high-stakes tests are non-standardized, such as a theater audition.

As with other tests, high-stakes tests may be criterion-referenced or norm-referenced. For example, a written driver's license examination typically is criterion-referenced, with an unlimited number of potential drivers able to pass if they correctly answer a certain percentage of questions. On the other hand, essay portions of some bar exams are often norm-referenced, with the worst essays failed and the best essays passed, without regard for the overall quality of the essays.

The "clear line" between passing and failing on an exam may be achieved through use of a cut score: for example, test takers correctly answering 75% or more of the questions pass the test; test takers correctly answering 74% or fewer fail, or don't "make the cut". In large-scale high-stakes testing, rigorous and expensive standard setting studies may be employed to determine the ideal cut score or to keep the test results consistent between groups taking the test at different times.

3.3 TEACHER MADE TESTS

Teacher made tests are classroom assessment normally prepared by teachers, have not been tested on sample populations of students and do not allow you to compare your students to a standard. Instead, these tests (also called criterion-reference tests), help us to test a student’s understanding of a particular (and often limited) body of knowledge. For example, if you are teaching a unit on ecology and want to determine whether your students have learned about predator-prey, you would include test questions about predator-prey relationships that related to
your specific objectives. You could also include questions dealing with knowledge and attitudes about predators and make the questions as easy or difficult as you wanted, based on the objectives you outlined earlier.

Many teachers prefer criterion-reference tests because the evaluation is based solely on the students' performance and the test relates directly to what was taught during the course. If everyone in the class can match the performance outlined in the objectives, then everyone gets top marks. Criterion-reference tests make a lot of sense for environmental education because if you design the right type of test, you can determine if your students have learned what you hoped they would learn. Criterion-reference tests can measure skills, knowledge and attitudes—the three major components of your environmental education program and they can be tailored to meet the environmental needs of your community.

The drawbacks to teacher-made tests are that they are often unreliable and less valid than standardized tests, and their effectiveness relies on the skill of the individual teachers who create the tests and grade them. When creating a test, it's important to match the test questions to your objectives. It is also helpful to ask yourself the following questions:

* Does this test match the level of difficulty I covered in the class?
* Does the test provide a representative sampling of the material I presented in class?
* Does each test question measure one or more of my objectives?

Of course, it's also important to match the type of test you give to the material you are trying to teach. For example, you might not want to give students an essay test to see if they learned key ecological vocabulary words, since a short answer test would be more efficient. But you might prefer an essay test if you are trying to evaluate how your students organize their thoughts and can analyze and evaluate an environmental problem in their community.

### 3.3.1 Characteristics of Teacher Made Tests

- Formative assessments give teachers a way to test knowledge and plan future instruction, but standardized exams or commercially prepared tests don't always accurately assess that information. The extra time required to prepare exams pays off with the potential for more accurate assessments, and with the benefits in mind, teachers can more accurately monitor student learning and progress.
- Teachers often expand on the textbook to make the information relevant and meaningful to their students. Instead of simply reading from the textbook, a
teacher might use non-fiction books, guest speakers, experiments, field trips and demonstrations to teach course content. Because tests provided with a textbook don't include the knowledge the students gain from outside experiences, teacher-made tests better reflect what is taught in class and fit better with the teaching methods they use. With customized tests, teachers can assess the students as they progress to check for understanding.

- Commercially prepared tests are typically multiple choice or fill-in-the-blank, although you may find some short answer and essay questions. When a teacher creates his or her own tests, she has complete control over the format. Paper-and-pencil tests can include different types of questions and formats that best fit specific material. Multiple-choice questions may work well for certain sections, while answering essay questions is best for others. Teachers also have the option of alternative testing types, such as an oral exams or presentations.

- Standardized testing typically only happens once per year, so the results don't necessarily give the teachers the tools to consistently improve teaching methods. Similarly, the tests provided by publishers to accompany textbooks are often only provided at the end of chapters or units. When a teacher makes her own exams, she can make as many or as few as she wants. More frequent testing gives a better look at the students' progress throughout the chapter and over the course of the year.

- The teacher knows ones students better than any publisher. While commercial tests may cover the majority of information, they may not take into account students with special needs or different learning styles. A teacher who makes her own tests has the option to tailor the exams to the students in her classroom. Examples include adding pictures or diagrams, enlarging print or leaving extra space between sentences to allow for easier reading.

3.4 STANDARDIZED TESTS

Standardized testing means that a test is “administered and scored in a predetermined, standard manner” (Popham, 1999). Students take the same test in the same conditions at the same time, if possible, so results can be attributed to student performance and not to differences in the administration or form of the test (Wilde, 2004). For this reason, the results of standardized tests can be compared across schools, districts or province. Standardized testing is also used as a shorthand expression for machine scored multiple-choice tests. As we will see, however, standardized tests can have almost any format. A standardized test is any form of test that requires all test takers to answer the same questions, or a selection of questions from common bank of questions, in the same
way, and that is scored in a “standard” or consistent manner, which makes it possible to compare the relative performance of individual students or groups of students. While different types of tests and assessments may be “standardized” in this way, the term is primarily associated with large-scale tests administered to large populations of students, such as a multiple-choice test given to all the eighth-grade public-school students in a particular district or province, for example.

In addition to the familiar multiple-choice format, standardized tests can include true-false questions, short-answer questions, essay questions or a mix of question types. While standardized tests were traditionally presented on paper and completed using pencils, and many still are, they are increasingly being administered on computers connected to online programs while standardized tests may come in a variety of forms, multiple-choice and true-false formats are widely used for large-scale testing situations because computers can score them quickly, consistently and inexpensively. In contrast, open-ended essay questions need to be scored by humans using a common set of guidelines or rubrics to promote consistent evaluations from essay to essay—a less efficient and more time-intensive and costly option that is also considered to be more subjective.

Many test experts and educators consider, standardized tests to be a fair and objective method of assessing the academic achievement of students, because of the standardized format, which reduces the potential for favoritism, bias or subjective evaluations. On the other hand, subjective human judgment enters into the testing process at various stages e.g., in the selection and presentation of questions, or in the subject matter and phrasing of both questions and answers. Subjectivity also enters into the process when test developers set passing scores a decision that can affect how many students pass or fail or how many achieve a level of performance considered to be “proficient.”

Standardized tests may be used for a wide variety of educational purposes. For example, they may be used to determine a young child’s readiness for kindergarten, identify students who need special-education services or specialized academic support, place students in different academic programs or course levels or award of diplomas and other educational certificates.

### 3.5 BENEFITS AND LIMITATIONS OF TESTS

There are advantages and limitations with each item format, and teachers should choose the format that best suits the purposes for assessment. We will discuss the advantages and limitations of teacher made and standardized/High Stake test one by one.
3.5.1 Advantages and Limitations of Teacher Made Tests

- If teachers have less time to score the assessments, selected-response questions are advantageous because they can be scored faster than constructed-response items.
- Selected-response items also are superior to constructed-response items if the goal is to measure basic levels of Bloom's Taxonomy, such as knowledge or comprehension.
- Students can respond more quickly to selected-response items, allowing the teacher to assess a broader range of objectives across a given timeframe.
- Selected-response items are also considered more objective than constructed response questions because the latter items require teachers to score the responses, introducing rater error to the scores. Because reliability is increased by having more items with less error, selected-response items tend to yield more consistent scores relative to constructed-response items.
- Selected-response items present both correct and incorrect options to students; those items are more prone to guessing than constructed-response items. The probability that students can guess correctly depends on the number of distracters for each question, the test-taking skills of the student, and the quality of the distracters.
- Constructed-response items also take less time to create, so if teachers have little time to construct an exam, they should consider including more of those items on the test.
- Crafting reasonable and high-quality distracters and selected-response items that are not prone to guessing is an arduous and time-consuming process.
- Constructed-response items are more suited for measuring more advanced levels of Bloom's Taxonomy in a direct manner. For example, if students are to demonstrate their evaluation skills or show that they can apply their knowledge in a novel situation, teachers must rely on constructed-response questions.
- Constructed-response items test the recall of information and actual demonstration of advanced skills, whereas selected-response items focus on mental recognition and serve, at best, as indirect indicators of advanced intellectual skills.

Activity 3.3 Discuss in a group the Limitation of Teacher Made Tests
3.5.2 Advantage and Disadvantage of High Stake Testing

High-stakes testing is one of the most controversial and contentious issues in education today, and the technicalities of the debates are both highly complex and continually evolving. The following arguments are commonly made for and against high-stakes testing.

Benefits of High Stake Testing
- It holds teachers accountable for ensuring that all students learn what they are expected to learn. While no single test can measure whether students have achieved all state learning standards (standardized tests can measure only a fraction of these standards), test scores are nevertheless one method used to determine whether students are learning at a high level.
- Motivates students to work harder, learn more, and take the tests more seriously, which can promote higher student achievement.
- Establishes high expectations for both educators and students, which can help reverse the cycles of low educational expectations, achievement, and attainment that have historically disadvantaged some student groups, particularly students of color, and that have characterized some schools in poorer communities or more troubled urban areas.
- Reveals areas of educational need that can be targeted for reform and improvement, such as programs for students who may be underperforming academically or being underserved by schools.
- Provides easily understandable information about school and student performance in the form of numerical test scores that reformers, educational leaders, elected officials and policy makers can use to develop new laws, regulations, and school-improvement strategies.
- Gives parents, employers, colleges and others more confidence that students are learning at a high level or that high school graduates have acquired the skills they will need to succeed in adulthood.

Disadvantage of High-Stakes Testing
- It forces educators to “teach to the test”—i.e., to focus instruction on the topics that are most likely to be tested, or to spend valuable instructional time prepping students for tests rather than teaching them knowledge and skills that may be more important.
- It promotes a more “narrow” academic program in schools, since administrators and teachers may neglect or reduce instruction in untested—
but still important—subject areas such as art, health, music, physical education or social studies, for example.

- It may contribute to higher, or even much higher, rates of cheating among educators and students, including coordinated, large-scale cheating schemes perpetrated school administrators and teachers who are looking to avoid the sanctions and punishments that result from poor test performance. Systematically changing test answers, displaying correct answers for students while they are taking the test, and disproportionately targeting historically low-performing students for expulsion are just a few examples taken from recent scandals.

- It has been correlated in some research studies to increase failure rates, lower graduation rates, and higher dropout rates, particularly for minority groups, students from low-income households, students with special needs and students with limited proficiency in English.

- May diminish the overall quality of teaching and learning for the same disadvantaged students who are the intended beneficiaries of high-stakes testing. Because of strong pressure on schools and teachers to improve test results and avoid score-based penalties, students of color and students from lower-income households and communities may be more likely to receive narrowly focused, test-preparation-heavy instruction instead of an engaging, challenging, well-rounded academic program.

- Exacerbates negative stereotypes about the intelligence and academic ability of minority students, who may worry so much about confirming negative racial stereotypes that they underperform on important exams (a phenomenon generally known as “stereotype threat”). And if such students perform so poorly that they fail a high-stakes graduation test, the failure will only limit their opportunities in higher education or future employment, which only perpetuates and reinforces the conditions that give rise to stereotypes.

Activity 3.4 Discuss with your tutorial group advantage of High stake Testing

3.6 CONCEPT OF USE OF TAXONOMIES IN TEST DEVELOPMENT

The taxonomies provide help not only in writing educational objectives but in writing assessment items also. Questions (items) on quizzes and exams demand different levels of thinking skills. For example, some questions might be simple
memorization of facts, and others might require the ability to synthesize information from several sources to select or construct a response. These taxonomies not only provide an ease to construct the test items appropriately and provide an opportunity to make objectives measurable and achievable. Taxonomies provide hierarchal matrix for the teachers to provide to proceed from lower level to higher level.

3.6.1 Using Bloom’s Taxonomy in Test Development

(a) What is Bloom's Taxonomy? Bloom's taxonomy is a classification system for the cognitive skills used in learning. Teachers use this taxonomy to plan lessons. A taxonomy is a system that groups and orders concepts or things, such as the classifications in biology that include family, genus and species. In 1956, Benjamin Bloom, an educational psychologist, created taxonomy of the cognitive skills required for learning.

(b) The Six Levels of Intellectual Skills Bloom's Taxonomy has six levels of intellectual skills, each one building on the previous level: knowledge, comprehension, application, analysis, synthesis and evaluation. This taxonomy is often represented by a pyramid divided into six sections. The bottom section is knowledge. At this level, children memorize facts and details. This is the foundation for all other cognitive skills and so most time is devoted to it in schools. The second level understanding. It is not enough to simply memorize facts and details; a child needs to understand the concepts. Once children understand concepts, they must be able to apply them in different situations. As we move up the pyramid, the cognitive skills required become more and more demanding. Analyzing requires students to consider the parts of something and think about what they mean. They may need to compare and contrast two things, for example. Synthesis requires that students go beyond the picture what they see or read. The last, top level, of the pyramid is evaluation. At this level, students work on forming an opinion and explaining the reasoning behind their opinion. Such opinions require that students have managed to move upwards through the levels from gaining knowledge all the way up to being able to make judgments.
<table>
<thead>
<tr>
<th>Bloom’s Level</th>
<th>Key Verbs (keywords)</th>
<th>Example Learning Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Creating</td>
<td>Design, formulate, build, invent, create, compose, generate, derive, modify and develop.</td>
<td>By the end of this lesson, the student will be able to determine whether using conservation of energy or conservation of momentum would be more appropriate for solving a dynamics problem.</td>
</tr>
<tr>
<td>Evaluating</td>
<td>Choose, support, relate, determine, defend, judge, grade, compare, contrast, argue, justify, support, convince, select and evaluate.</td>
<td>By the end of this lesson, the student will be able to design an original homework problem dealing with the principle of conservation of energy.”</td>
</tr>
<tr>
<td>Analyzing</td>
<td>Classify, break down, categorize, analyze, diagram, illustrate, criticize, simplify and associate.</td>
<td>By the end of this lesson, the student will be able to differentiate between potential and kinetic energy.</td>
</tr>
<tr>
<td>Applying</td>
<td>Calculate, predict, apply, solve, illustrate, use, demonstrate, determine, model, perform and present.</td>
<td>By the end of this lesson, the student will be able to calculate the kinetic energy of a projectile.</td>
</tr>
<tr>
<td>Understanding</td>
<td>Describe, explain, paraphrase, restate, give original examples of, summarize, contrast, interpret and discuss.</td>
<td>By the end of this lesson, the student will be able to describe Newton’s three laws of motion to in her/his own words</td>
</tr>
<tr>
<td>Remembering</td>
<td>List, recite, outline, define, name, match, quote, recall, identify, label and recognize.</td>
<td>By the end of this lesson, the student will be able to recite Newton’s three laws of motion.</td>
</tr>
</tbody>
</table>

(c) Revision of Bloom's Taxonomy
In the 1990's, the taxonomy was revised, replacing the nouns with verbs. Instead of knowledge, comprehension, application, analysis, synthesis and evaluation, the revised version lists remembering, understanding, applying, analyzing, evaluating and creating. Evaluating is no longer the top level. It replaces synthesis and then creating goes at the top. Technically, though synthesizing and evaluating have just switched places. The idea behind the switch is that before someone can create something new - synthesize - he
has to be able to evaluate the information he already has. Creating or synthesizing is considered the most difficult mental skill.

(d) Utilizing Bloom’s Taxonomy for Assessment

Bloom’s Taxonomy provides useful structure to help us categorize test question for assessment of students’ learning. The teacher may plan a test, which includes a variety of questions, forcing the students to think and function at each level of the taxonomy. This is the teacher’s opportunity to be creative. The teacher prepares questions related to all levels of the taxonomy directly related to the content of study. Below is an example of some questions and activities related to the study of domesticated animals utilizing the Bloom’s Taxonomy methodology? Notice the use of the verbs in each question.

Example

Topic: Domesticated Animals

Level I and II: Knowledge/Comprehension

Questions
1. Locate and list the animals that live on your block.
2. Identify the different breeds of dogs in your neighborhood.
3. Observe a dog while it is at play and rest. Explain how different dogs sit and lay.

Level III: Application
1. Teach your dog a new trick.
2. Interview people who own pets. Make a survey of people who own pets in your neighborhood.
3. Construct a mobile or draw a collage about dog care and grooming.

Level IV: Analysis
1. Compare and contrast the physical and social characteristics of dogs and cats.
2. Develop a survey comparing and contrasting the different types of foods available for dogs or cats.
3. Make a chart comparing the anatomy of dogs and cats.

Level V: Synthesis
1. Develop a cartoon based on the relationship between an animal and a child.
2. Invent a toy or machine that would help dogs or cats live a healthier and happier life. 3. Create a TV game show about domesticated animals.

**Level VI: Evaluation**
1. Lead a panel discussion on the values of pets.
2. Write an editorial about the advantages and disadvantages of having a pet animal.
3. Have a dog and cat show. Present winner awards and ribbons.

**Self Assessment Task 3.2: Construct a test consisting two test Items of Subject of Your choice of 8th Standard Covering all the Taxonomical Level of Bloom’s Taxonomy of Cognitive Domain**

**3.6.2 Using SOLO Taxonomy in Test Development**

“SoLO Taxonomy provides a simple and robust way of describing how learning outcomes grow in complexity from surface to deep understanding” (Biggs & Collis)

**What is SOLO Taxonomy?**

SOLO (Structure of Observed Learning Outcomes) provides a structured framework for students to use to progress their thinking and learning. It encourages students to think about where they are currently with their learning and what they need to do in order to progress. There are five main stages:

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prestructural</td>
<td>This is the first stage – where students don’t really have any knowledge or understanding of the topic being studied. A student who is pre-structural will usually respond with ‘I don’t understand’.</td>
</tr>
<tr>
<td>Unistructural</td>
<td>Moving on from pre-structural, students who are uni-structural have a limited knowledge of the topic – they may just know one isolated fact about the topic. So, a typical response might be: ‘I have some understanding of this topic’</td>
</tr>
</tbody>
</table>
Progressing from uni-structural to multi-structural simply means that the student knows a few facts about this topic – but is unable to link them together. So a typical response might be ‘I know a few things about this topic’ or ‘I have gathered some information about this topic’.

With relational, we are starting to move towards higher level thinking – students are able to link together and explain several ideas around a related topic. So a typical student ‘relational response might be: ‘I can see the connections between the information I have gathered’.

The final and most complex level is extended abstract. With this, not only students are able to link lots of related ideas together, but they can also be link these to other bigger ideas and concepts. So a student response at this level might sound like: ‘By reflecting and evaluating on my learning, I am able to look at the bigger picture and link lots of different ideas together’.

An example….

Using SOLO Taxonomy for Development of a Biology Test

In science, students might be asked the question ‘What do you understand by the term respiration’. Students may then respond in the following ways:

- **Prestructural** – “Err…..What?”
- **Unistructural** – “It releases energy“
- **Multistructural** – “It’s a chemical reaction that releases energy, uses oxygen and glucose and release carbon dioxide.”
- **Relational** – “It’s a reaction that takes place in all body cells. Products of digestion, such as glucose, are transported to cells by the blood and reacted with oxygen to produce carbon dioxide – which is breathed out. Energy is released.”
- **Extended abstract** – “It’s a reaction that takes place in all body cells. Products of digestion, such as glucose, are transported to cells by the blood and reacted with oxygen to produce carbon dioxide – which is breathed out via the lungs (using gas exchange and ventilation). As energy is released,
respiration is an example of an exothermic reaction. The energy that is released can then be used by the body for growth of new cells, repair of tissues and keeping warm.”

**Why is it so useful?**

- It supports students to reflect on their own thinking
- It helps teachers to thoughtfully shape learning intentions and learning experiences.
- It makes it easy to identify and use effective success criteria.
- It provides feedback and feed forward with regards to learning outcomes.
- It helps students to reflect meaningfully on what the next steps in their learning are.

### 3.7 PROCEDURE OR STEPS FOR A STANDARDIZED TEST DEVELOPMENT PROCESS

Standardized tests offer a uniform test-taking experience. They are set up in such a way that the test conditions and scoring have a specific procedure that is interpreted in a consistent manner from one administration to the next. Standardized tests are created by test development specialists who ensure that the examinations have a specific goal, specific intent and solid written foundation. When creating a standardized test, developers follow a certain set of steps to ensure quality.

**Purpose**

The first step in standardized test development is establishing a purpose. Developers consider how the test scores will be used, whether for certification, licensing, minimum competency or mastery of a specific subject. The purpose determines both the length and difficulty of the exam. Taking time in the beginning of the process to develop a clear purpose ensures that "your goals and priorities will be more effectively met,"

**Specifications**

Standardized test specifications are composed of a description and a blueprint. The description outlines the test's purpose and the targeted examinee population. A test blueprint is developed, providing a list of content area and cognitive levels that will be tested. The blueprint also specifies the amount of content each area of the test will contain.
Development
An initial pool of items is developed during the writing and development phase. A panel of subject matter experts is formed to write each section of test items. Each item is written based on the specifications in the blueprint.

Review
After the test items are written, they are reviewed. The review addresses potential problems and ensures that the content is clear, accurate and free of technical flaws. The test must prove that it is fair for all examinees and subgroups.

Pilot
When the initial development and review phases are complete, the pilot exam must go through a field test in which an appropriate group of examinees takes the test. This ensures the test's quality and practical application. The test is scored; however, the pilot is not used to measure examinees' knowledge.

Forms, Scoring and Analysis
The final steps in creating a standardized test are assembling the test forms, determining a passing score and item analysis. During the form assembly phase, multiple test forms are developed for each test as a security measure. Using statistical methods and policy, a passing score is then developed for the test. Item analysis is one of the most important steps in the test development process. It determines whether the test items are too simple, too difficult and display the skill level and knowledge of the examinee.

3.8 EXAMPLES OF STANDARDIZED TESTS WITH CHARACTERISTICS
The Standardized tests can be classified as per their functions are

Group and Individual Tests
These are used to screen large groups of people to identify possible candidates for individual testing. Group tests are, compared to individual ones, inexpensive with respect to time and money; the trade-off for being less expensive is that they do not provide a great deal of information. Individual tests provide detailed information about the individual but cost a great deal. Financially, schools cannot afford to test all students individually, and it is unethical to collect the type of
detailed, private information on students unless there is a professionally sound reason. Individual tests should only be used when there is a need (based on the results of a group test, parent/teacher referral, etc.); unless such a need exists, the use of an individual (or group) test may constitute an unwarranted invasion of privacy.

**Norm-referenced**

These tests compare one's abilities to the other individuals taking the test (currently or in the original norming group).

**Achievement Tests**

These tests measure what one has learned in a particular area (such as mathematics) that has been directly taught in a curriculum; they are often criterion-referenced.

**Criterion-referenced**

These tests compare the individual's abilities against some established standard (not just the norming group's performance).

**Aptitude**

Achievement Aptitude tests are used for selection and to predict future performance in an area (an occupation, an educational program, etc.); they generally are norm referenced and often do not cover information that was directly taught in a curriculum. There are several types of aptitude tests.

**Personality**

These tests describe the characteristic ways one behaves,

**Projective**

These tests are used to help diagnose emotional disturbance, typically by presenting an ambiguous stimulus that the taste must interpret. How people interpret the stimulus reflects their underlying personality characteristics and pathology because they "project" these onto the stimulus

**Interest Inventories**

These compare one's expressed likes/dislikes with those of successful people in various occupations; they are used extensively in career planning
Intelligence tests (sometimes considered a variation of achievement instruments) These are very narrowly attempt to predict one's ability to learn new materials (based on what one has already learned-achievement, and how quickly one is able to learn).

Characteristics of Standardize Tests

Standardized tests are evaluated on the basis of their validity and reliability. Validity, very generally is concerned with how well a test measures what it is supposed to (e.g., accuracy), while reliability is the extent to which a test consistently measures what it is measuring. (Notice that a test can be reliable but not valid!) There are several types of validity: Content (measures the degree of relationship between what's taught and what's tested), construct (does the test measure the particular psychological domain-e.g., IQ-that it purports to) and criterion-related.

Criterion-related validity attempts to show that one's performance on a particular instrument/test is correlated with one's performance on some external standard right now or in the future (i.e., predictive). For example, if one scored high on a maths achievement test, it would be reasonable to expect that that person would be earning a high grade in maths (e.g., the student's performance in the maths class is concurrently validating the score on the test). Predictive validity also attempts to relate test performance to an external standard of performance, but the external standard is temporally distant in that one's performance on today's test is used to predict how one will perform on the external standard at some point in the future (predictive).

Interpreting Test Results. Tests are scored by hand or computer and a profile (or report if it is an individual test) is prepared for each student. Performance in each area and sub-area of the test is presented as both a raw score (which is usually meaningless unless one also has -and knows how to use- the test's measures of central tendency and variability) and a derived scored (a raw score that has been transformed/converted so it has normative meaning by itself [many computer-scored tests also provide graphs and charts of the students' scores to aid in their interpretation])

3.8.1 Reliability of Tests

Reliability in statistics and psychometrics is the overall consistency of a measure. A measure is said to have a high reliability if it produces similar results under consistent conditions. For example, measurements of people's height and weight are often extremely reliable. Reliability is the degree to which an assessment tool produces stable and consistent results. A reliable test is one that will give the
same results over and over again. It's consistent, dependable and stable. It's important that a test is reliable so that you can count on the results. For example, if you give the same test to the same group of students three times in a row in a short period of time, the results should not fluctuate widely. If you use a different form of the test, the results should also remain consistent. If they don't, the test is not reliable. For example, if you have two test items to measure one objective, do the students who get one right also get the other right and the students who get one wrong get the other one wrong too? You want a test to be reliable so that you can count on it to test for the same things no matter who you give it to and when you give it. To improve reliability, you can increase the number of test items, give the test to a mixed student group, include test items that are of moderate difficulty rather than of mainly easy or hard questions, double check to make sure all test items are clear and understandable and use test items that can be scored objectively rather than subjectively.

### 3.8.1.1 Types of Reliability

(a) **Test-retest reliability** is a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals. The scores from Time 1 and Time 2 can then be correlated in order to evaluate the test for stability over time. *Example:* A test designed to assess student learning in psychology could be given to a group of students twice, with the second administration perhaps coming a week after the first. The obtained correlation coefficient would indicate the stability of the scores.

(b) **Parallel forms reliability** is a measure of reliability obtained by administering different versions of an assessment tool (both versions must contain items that probe the same construct, skill, knowledge base, etc.) to the same group of individuals. The scores from the two versions can then be correlated in order to evaluate the consistency of results across alternate versions. *Example:* If you wanted to evaluate the reliability of a critical thinking assessment, you might create a large set of items that all pertain to critical thinking and then randomly split the questions up into two sets, which would represent the parallel forms.

(c) **Inter-rater reliability** is a measure of reliability used to assess the degree to which different judges or raters agree in their assessment decisions. Interrater reliability is useful because human observers will not necessarily interpret answers the same way; raters may disagree as to how well certain responses or material demonstrate knowledge of the construct or skill being assessed. *Example:* Inter-rater reliability might be employed when different judges are evaluating the degree to which art portfolios meet certain standards. Inter-rater reliability is especially useful when judgments can be
considered relatively subjective. Thus, the use of this type of reliability would probably be more likely when evaluating artwork as opposed to maths problems.

(d) **Internal consistency reliability** is a measure of reliability used to evaluate the degree to which different test items that probe the same construct produce similar results.

(e) **Average inter-item correlation** is a subtype of internal consistency reliability. It is obtained by taking all of the items on a test that probe the same construct (e.g., reading comprehension), determining the correlation coefficient for each pair of items, and finally taking the average of all of these correlation coefficients. This final step yields the average inter-item correlation.

(f) **Split-half reliability** is another subtype of internal consistency reliability. The process of obtaining split-half reliability is begun by “splitting in half” all items of a test that are intended to probe the same area of knowledge (e.g., World War II) in order to form two “sets” of items. The entire test is administered to a group of individuals, the total score for each “set” is computed and finally the split-half reliability is obtained by determining the correlation between the two total “set” scores.

### 3.8.2 Validity of Tests

It refers to how well a test measures what it is purported to measure. When a test is valid, it measures what it's designed to measure. For example, if you are trying to test if your students have achieved the following objective "Given a plow, students will be able to plow on the contour to help prevent soil erosion" but test by using a test item that asks why it's important to plow on the contour, your test will not provide a valid measure of this objective. To test for that objective, you need to actually see the students plow. Or if your objective is to have students’ list three causes of reef destruction, but the test question has students’ list three causes of ocean pollution, the test item doesn't match the objective. If the test question were to ask students to list three causes of reef destruction, the question would be valid.

One way to make sure your test is valid is to double check each test item and make sure each is measuring your pre-determined objectives. You can also ask your colleagues to rate your questions against your objectives to make sure they match.

#### 3.8.2.1 Types of Validity

(a) **Face Validity** ascertains that the measure appears to be assessing the intended construct under study. The stakeholders can easily assess face validity. Although this is not a very “scientific” type of validity, it may be
an essential component in enlisting motivation of stakeholders. If the stakeholders do not believe the measure is an accurate assessment of the ability, they may become disengaged with the task.

Example: If a measure of art appreciation is created all of the items should be related to the different components and types of art. If the questions are regarding historical time periods, with no reference to any artistic movement, stakeholders may not be motivated to give their best effort or invest in this measure because they do not believe it is a true assessment of art appreciation.

(b) **Construct Validity** is used to ensure that the measure is actually measure what it is intended to measure (i.e. the construct), and not other variables. Using a panel of “experts” familiar with the construct is a way in which this type of validity can be assessed. The experts can examine the items and decide what that specific item is intended to measure. Students can be involved in this process to obtain their feedback. Example: A women’s studies program may design a cumulative assessment of learning throughout the major. The questions are written with complicated wording and phrasing. This can cause the test inadvertently becoming a test of reading comprehension, rather than a test of women’s studies. It is important that the measure is actually assessing the intended construct, rather than an extraneous factor.

(c) **Criterion-Related Validity** is used to predict future or current performance - it correlates test results with another criterion of interest. Example: If a physics program designed a measure to assess cumulative student learning throughout the major. The new measure could be correlated with a standardized measure of ability in this discipline, such as an ETS field test or the GRE subject test. The higher the correlation between the established measure and new measure, the more faith stakeholders can have in the new assessment tool.

(d) **Formative Validity** when applied to outcomes assessment it is used to assess how well a measure is able to provide information to help improve the program under study. Example: When designing a rubric for history one could assess student’s knowledge across the discipline. If the measure can provide information that students are lacking knowledge in a certain area, for instance the Civil Rights Movement, then that assessment tool is providing meaningful information that can be used to improve the course or program requirements.

(e) **Sampling Validity** (similar to content validity) ensures that the measure covers the broad range of areas within the concept under study. Not
everything can be covered, so items need to be sampled from all of the domains. This may need to be completed using a panel of “experts” to ensure that the content area is adequately sampled. Additionally, a panel can help limit “expert” bias (i.e. a test reflecting what an individual personally feels are the most important or relevant areas). Example: When designing an assessment of learning in the theatre department, it would not be sufficient to only cover issues related to acting. Other areas of theatre such as lighting, sound, functions of stage managers should all be included. The assessment should reflect the content area in its entirety.

3.8.3 Usability of Tests

Usability testing refers to evaluating a product or service by testing it with representative users. Typically, during a test, participants will try to complete typical tasks while observers watch, listen and takes notes. You should also select tests based on how easy the test is to use. In addition to reliability and validity, you need to think about how much time you have to create a test, grade it and administer it. You need to think about how you will interpret and use the scores from the tests. And you need to check to make sure the test questions and directions are written clearly, the test itself is short enough not to overwhelm the students, the questions don’t include stereotypes or personal biases, and that they are interesting and make the students think.

3.9 CURRENT TRENDS IN HIGH STAKE TESTING

Following are the trends in the High Stake testing in educational field:

(a) **High Stake Tests are used to hold schools and educators accountable for educational results and student performance.** In this case, test scores are used as a measure of effectiveness, and low scores may trigger a variety of consequences for schools and teachers.

(b) **High Stake Tests are used to evaluate whether students have learned what they are expected to learn,** such as whether they have met state learning standards. In this case, test scores are seen as a representative indicator of student achievement.

(c) **High Stake Tests are used to identify gaps in student learning and academic progress.** In this case, test scores may be used, along with other information about students, to diagnose learning needs so that educators can provide appropriate services, instruction or academic support.

(d) **They are used to identify achievement gaps among different student groups,** including students of color, students who are not proficient in
English, students from low-income households and students with physical or learning disabilities. In this case, exposing and highlighting achievement gaps may be seen as an essential first step in the effort to educate all students well, which can lead to greater public awareness and changes in educational policies and programs.

(e) **They are used to determine whether educational policies are working as intended.** In this case, elected officials and education policy makers may rely on standardized-test results to determine whether their laws and policies are working or not, or to compare educational performance from school to school or state to state. They may also use the results to persuade the public and other elected officials that their policies are in the best interest of children and society.

(f) The field of High stake testing has evoked many controversies like: are numerical scores on a standardized test misleading indicators of student learning, since standardized tests can only evaluate a narrow range of achievement using inherently limited methods? Or do the scores provide accurate, objective and useful evidence of school, teacher or student performance? (Standardized tests don’t measure everything students are expected to learn in school. A test with 50 multiple-choice questions, for example, can’t possibly measure all the knowledge and skills a student was taught or is expected to learn, in a particular subject area, which is one reason why some educators and experts caution against using standardized-test scores as the only indicator of educational performance and success.)

(g) Are standardized tests fair to all students because every student takes the same test and is evaluated in the same way? Do the tests have inherent biases like language barriers etc?

### 3.10 Summary of the Unit

Classroom assessment is the process, usually conducted by teachers, of designing, collecting, interpreting and applying information about student learning and attainment to make educational decisions. There are four interrelated steps to the classroom assessment process. These are defining the purpose of assessment, measuring student learning or attainment, evaluating the measured data and finally the interpretations to guide instruction, render grades or help students with any particular learning. Classroom Assessment are teacher made tests which are aimed to measure academic achievement, identify learning problems or inform instructional adjustments, among other purposes that is why they are also called “teacher made tests or low stake tests” A high-stakes test is a single, defined assessment; that has as a clear line drawn between those who pass and those who fail; and has direct consequences for passing or failing (something "at stake").
Teacher made tests are important for giving formative information during learning to provide intervention required; covers whole contents; used more frequently and caters for all type of students. Standardized tests are tested on large group and time tested to provide valid and reliable information. A standardized test is any form of test that requires all test takers to answer the same questions or a selection of questions from common bank of questions are scored in a “standard” or consistent manner, which makes it possible to compare the relative performance of individual students or groups of students. While different types of tests and assessments may be “standardized” High stake tests are used to make the stakeholders accountable, while teacher made tests are used to provide during instruction feedback. Taxonomies help us to write the educational objectives according to levels of increasing difficulty and help us in assessment. Such as Bloom's Taxonomy has six levels of cognitive skills, each one building on the previous level: knowledge, comprehension, application, analysis, synthesis and evaluation. And SOLO Taxonomy has five levels prestructural, unistructural, multistructural, relational and extended abstract. High stake testing is not new in advanced countries but new in Pakistan. There are many controversies related to it because of their accountability and decision based on standardized testing but it on the other prohibits many unprofessional individuals for joining any position or profession.

3.11 SELF ASSESSMENT QUESTIONS

Q.1 Teacher Made tests are more fruitful than standardized tests for teachers. Elucidate with argument.
Q.2 High Stake Tests are also authentic assessments. Comment the statement.
Q.3 Construct a test consisting two test items of subject of your choice of 8th standard covering all the Taxonomical Level of Bloom’s Taxonomy of Cognitive Domain.
Q.4 Construct test containing two test items each of subject of your choice of 9th standard covering all the Taxonomical Level of SOLO Taxonomy.
Q.5 Enlist and explain the procedure for construction of standardized tests.
Q.6 What are the basic characteristics which make an assessment authentic? Discuss in detail.
BIBLIOGRAPHY


Basic Principles of Curriculum and Instruction. Chicago: University of Chicago Press. Copyright 2003-2009 The Gale Group, Inc. All rights reserved.
DEVELOPMENT OF SUPPLY TEST ITEM

Written By: Dr. Muhammad Azeem
Reviewed By: Dr. Muhammad Zafar Iqbal
CONTENTS

Introduction.................................................................................................................................71
Objectives .................................................................................................................................71
4.1 Determining the Behaviors to be Assessed.................................................................72
4.2 Developing Norms of the Test.......................................................................................74
4.3 Planning the Test..............................................................................................................75
4.4 Ensuring Content Validity (Course Coverage, Concept Coverage and Learning Outcomes Coverage) ..........................................................................................75
4.5 Constructing a Table of Specifications Based on Bloom’s Taxonomy .......76
4.6 Constructing a Table of Specifications Based on SOLO Taxonomy ........85
4.7 Writing Test Items (Supply and Selection Types) Based upon table of Specifications .................................................................................................................................87
  4.7.1 Fill in the Blanks..........................................................................................................87
  4.7.2 Short Answer Questions and Marking Guide .........................................................88
  4.7.3 Essay Type Questions and Rubric for Marking ...............................................91
4.8 Self Assessment Questions .............................................................................................94
Bibliography .............................................................................................................................95
INTRODUCTION

Supply test items require students to supply or construct their own responses and answers. These items are also referred to as constructed response question CRQs. If CRQs require response in terms of few words/numbers or very limited description, then these are called restricted response item RRI, otherwise extended response item ERI. Supply test items force the student to demonstrate mastery or understanding by supplying the evaluator with something. Whether it be a short answer or an hour long presentation, they must construct their own product or answer. For this reason, they are often used as formative assessments to monitor student learning and teaching effectiveness. For instance, using supply response assessment, a teacher can stop in the middle of instruction and ask students to summarize what his/her just taught. It gives his/her an on-the-spot-opportunity to evaluates her students' understanding by listening to their supplied responses. If they get it, he/she moves on otherwise may loops back around.

OBJECTIVES

After studying this unit, the students will be able to:
1. highlight the role of assessment in teaching and learning process.
2. discuss factors affecting the selection of subjective and objective types of questions for the classroom tests.
3. describe the various types of reporting test score
4. define an objective and an outcome
5. enlist the different types of techniques and explain their role in education system.
4.1 DETERMINING THE BEHAVIORS TO BE ASSESSED

A functional behavior assessment is a comprehensive and individualized strategy to identify the purpose or function of a student’s problem behavior(s).

- Develop and implement a plan to modify variables that maintain the problem behavior.
- Teach appropriate replacement behaviors using positive interventions.

Things to Do
- Define the problem behavior.
- Devise a plan to collect data.
- Compare and analyze the data.
- Formulate the hypothesis.
- Develop and implement a behavior intervention plan.
- Monitor the plan.

i. **Define the problem behavior**
   A well-defined behavior is essential when collecting data and communicating results with other team members. Define the behavior in specific, observable and measurable terms. In other words, something you can see and count.

ii. **Devise a plan to collect data**
   There are two basic methods for collecting data—direct and indirect. Indirect methods often use student records, interviews, questionnaires or checklists to identify how others perceive the situation and possible motivations for the problem behavior. Then more direct methods, such as observations, record the situational factors surrounding the problem behavior.

   - **Indirect Method**
     Begin by reviewing the student’s records. In a systematic fashion, identify any previous relevant background data from existing documents that relate to the problem behavior. When conducting an interview, consider asking the following questions:
     - . . . is present when the problem occurs?
     - . . . is happening just before the problem occurs and what happens immediately after the problem behavior?
     - . . . does the problem behavior occur?
     - . . . does the problem behavior take place?

   And finally, are there times or places when the problem behavior does NOT occur? Such information is valuable when planning future interventions. When
using indirect methods, watch for factors that can influence a student’s behavior, such as:

- Physiological
- Environmental
- Curricular and instructional
- Setting events or incidents that happen some-time before the problem situation.

Using a variety of indirect methods like these are especially important if there’s more than one reason for the student’s noncompliant behavior. To know if you’re on the right track, gather information from as many sources and places as possible, such as daily classes, counselors and teachers and after-school activities.

Direct Method

Next use direct assessment to observe and record the problem events as they happen. Direct assessments may include frequency counts, interval recording systems and antecedent-behavior-consequence or A-B-C charts. To use an A-B-C form start by writing the basic information, such as the student’s name, observer’s name, and so on. Then each time the student behaves inappropriately, record the behavior and what events occur just before. These events are called “antecedents.” Also be sure to record the consequences or what happens right after the incident.

Remember to record only those things you see or hear, not your interpretation of the behavior. Continue collecting data in this manner until a pattern emerges that shows a relationship between the student’s behavior and his environment. Typically this requires 2 to 5 days, depending on how often the behavior occurs and the frequency of the observation. The more data you collect, the more accurate the picture of the student’s day-to-day behavior and the events that surround it.

iii. Compare and analyze the data

To see the big picture, summarize the data. A summary could identify:

- Setting events
- Antecedents or triggering events
- The target behavior
- Consequences that maintain the behavior
Another summary could simply identify who, what, when and where. Organizing your data in these ways will help clarify relationships between the student’s behavior and related events or factors.

iv. **Formulate the hypotheses**
Based on the data you collect, give your best, educated guess to explain the function or reason for the behavior. Generally speaking, problem behaviors serve two basic functions:
- To get something
- To avoid and escape something

v. **Develop and Implement a Behavior Intervention Plan**
Students respond best to Behavior Intervention Plans (BIP) that use positive methods to encourage and teach appropriate alternative behaviors. For example, positive methods may include:
- Modifying the physical environment
- Adjusting the curriculum or instructional strategy
- Changing the antecedents or consequences for the student’s behavior
- Finally, teaching a more acceptable replacement behavior that serves the same function.

vi. **Monitor the plan**
Regardless of the behavior intervention plan that your team develops, be sure to regularly monitor the student’s progress over time. This means:
- Collect data on student progress
- Review and evaluate the behavior goals
- Then determine whether to continue or modify the BIP.

Setting review dates ensures that this will happen.

A FBA is not a one-shot process. You can do an FBA anytime you need useful information to design effective interventions, because conducting an FBA is not only complying with federal law, its basic good practice.

### 4.2 DEVELOPING NORMS OF THE TEST

**Norms** – This refers to the comparison of a student’s score in a test to the scores of a reference group of students. A norm that can be followed with confidence results in a good comparison. On the other hand, if the norms of a verbal ability test were based on a reference group composed of native English speakers and if the examinee’s English is his second language, his score could not be compared to the norms established by the test.

A test norm is a set of scalar data describing the performance of a large number of people on that test. Test norms can be represented by two important statistics:
Means and Standard Deviations. The most important measure in psychometrics is the arithmetical average or the mean. This is intuitive for many people. For example the average of 2, 4 and 6 is 4. Or the average of 1, 2, and 3 is 2. In psychometrics this average is important because it indicates the middle position in a distribution.

4.3 PLANNING THE TEST

1. Review Curriculum
2. Review Textbook or Learning Material
3. Compatibility between Curriculum and Textbook (Weightage: see outlines & then see textbook)
4. Decide Categories / Types of Test Items (MCQs, SAQs, ETQs) (NRT, CRT)
5. Decide Weightage of Different Test Items and Cognitive Abilities
6. Draw Table of Specification also called Test Specification and Grid Specification
7. Develop Questions according to Test Specification / TOS / GS
8. Review Questions (improve)
9. Piloting the Test:
   1. Difficulty Level (0 to 1) (Should be 0.27- 0.84)
   2. Discrimination Index (0 to 1) may be negative (should be 0.5 to onward)
   3. Power of Distracters (how a option is weak and powerful)
10. Finalizing the Test

4.4 ENSURING CONTENT VALIDITY (COURSE COVERAGE, CONCEPT COVERAGE AND LEARNING OUTCOMES COVERAGE)

Content validity, sometimes called logical or rational validity, is the estimate of how much a measure represents every single element of a construct.

For example, an educational test with strong content validity will represent the subjects actually taught to students, rather than asking unrelated questions.

Content validity is often seen as a prerequisite to criterion validity, because it is a good indicator of whether the desired trait is measured. If elements of the test are irrelevant to the main construct, then they are measuring something else completely, creating potential bias. In addition, criterion validity derives quantitative correlations from test scores. Content validity is qualitative in nature, and asks whether a specific element enhances or detracts from a test or research program.
i. **How is content validity measured?**
Content validity is related to face validity, but differs wildly in how it is evaluated. Face validity requires a personal judgment, such as asking participants whether they thought that a test was well constructed and useful. Content validity arrives at the same answers, but uses an approach based in statistics, ensuring that it is regarded as a strong type of validity. For surveys and tests, each question is given to a panel of expert analysts, and they rate it. They give their opinion about whether the question is essential, useful or irrelevant to measuring the construct under study. Their results are statistically analyzed and the test modified to improve the rational validity.

ii. **An example of low content validity**
Let us look at an example from employment, where content validity is often used. A school wants to hire a new science teacher and a panel of governors begins to look through the various candidates. They draw up a shortlist and then set a test, picking the candidate with the best score. Sadly, he proves to be an extremely poor science teacher. After looking at the test, the education board begins to see where they went wrong. The vast majority of the questions were about physics so, of course, the school found the most talented physics teacher. However, this particular job expected the science teacher to teach biology, chemistry and psychology. The content validity of test was poor and did not fully represent the construct of 'being a good science teacher.' Suitably embarrassed, the school redesigned the test and submitted it to a panel of educational experts. After asking the candidates to sit the revised test, the school found another teacher and she proved to be an excellent and well-rounded science teacher. This test had a much higher rational/content validity and fully represented every element of the construct.

### 4.5 CONSTRUCTING A TABLE OF SPECIFICATION BASED ON BLOOM’S TAXONOMY

A table of specification (TOS) is the scientific expression specified to the plan for writing test items. A table of specification reveals what has been taught in the teaching and learning process. Infact, table of specification is a mirror of mode of two dimensions’ instructional process like content and intellectual procedure. At the same time table of specification development and usage are continuing theme to discussion.

i. **Alignment of Student Learning Outcomes SLOs and Bloom’s Taxonomy**
Action verbs of SLOs should be aligned with careful review. This alignment provides basis for development of table of specification. Curricula may or may not show this alignment. If there exist already this alignment in curriculum, then OK otherwise test developers have to do this.
Below is the sample of this alignment.

<table>
<thead>
<tr>
<th>Sr No</th>
<th>Content</th>
<th>Sub Content</th>
<th>SLOs No</th>
<th>SLO</th>
<th>Cognitive level</th>
<th>Que type</th>
<th>No of Que in each strand and its detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reading and thinking skills</td>
<td>1</td>
<td>*Recognize each paragraph as a separate meaningful unit of expression.</td>
<td>KNW</td>
<td>MCQ/RRI</td>
<td>Total Que = 11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.1</td>
<td>The main idea in a paragraph is carried in a sentence, called a topic sentence.</td>
<td>KNW</td>
<td>MCQ/RRI</td>
<td>MCQs= 10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>other sentence in the paragraph support the topic sentence.</td>
<td>KNW</td>
<td>MCQ/RRI</td>
<td>Marks of MCQs= 20</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2</td>
<td>*Recognise specific parts of words including Common inflectional endings and Compound words.</td>
<td>KNW</td>
<td>MCQ</td>
<td>No of OEQ= 1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Writing Skills</td>
<td>12</td>
<td>*Write multi-syllable words with correct spellings.</td>
<td>KNW</td>
<td>MCQ/SAQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>13</td>
<td>*Write sentences of their own using correct capitalization, punctuation and spellings.</td>
<td>COM</td>
<td>MCQ/SAQ</td>
<td>Total Que= 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>*Write a guided paragraph using ideas gathered and organized through various strategies.</td>
<td>SYN</td>
<td>OEQ</td>
<td>No of OEQ= 5</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Oral Communication</td>
<td>22</td>
<td>Identify and use previously learnt and more formulaic expressions for greetings and routine social courtesies according to the age, gender and status of addressee.</td>
<td>COM</td>
<td>MCQ/RRI</td>
<td>Total MCQs= 12</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Formal and Lexical Aspects of Language</td>
<td>23</td>
<td>*Recognize, find out, create and use more rhyming words.</td>
<td>COM</td>
<td>MCQ</td>
<td>Marks of MCQs= 24</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>23.1</td>
<td>*Locate identify, and differentiate between and use some simple pair of words including homophones, homonyms</td>
<td>COM</td>
<td>MCQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Grammar and Structure Nouns</td>
<td>24</td>
<td>*Recall, and demonstrate use of more common, countable and uncountable, collective nouns from immediate and extended environment.</td>
<td>COM</td>
<td>MCQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24.1</td>
<td>*Recognize and use nouns with no change in number.</td>
<td>KNW</td>
<td>MCQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24.2</td>
<td>Classify more nouns as common and proper nouns (names of people, pets, places, mountains, lakes, rivers, etc.).</td>
<td>COM</td>
<td>MCQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pro nouns</td>
<td>25</td>
<td>Illustrate use of pronouns learnt earlier. Use the personal pronouns myself, yourself/ves, himself, herself, ourselves, themselves and itself.</td>
<td>COM</td>
<td>MCQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sr No</td>
<td>Content</td>
<td>Sub Content</td>
<td>SLOs No</td>
<td>SLO</td>
<td>Cognitive level</td>
<td>Que type</td>
<td>No of Que in each strand and its detail</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-------------</td>
<td>---------</td>
<td>-----</td>
<td>----------------</td>
<td>---------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25.1</td>
<td>*Show possession by using the pronouns my, your, his, her, its, our, and their before nouns.</td>
<td>COM</td>
<td>MCQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25.2</td>
<td>*Recognize that pronouns agree with their nouns in gender and number.</td>
<td>KNW</td>
<td>MCQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25.3</td>
<td>Identify and illustrate extended use of words that point to something.</td>
<td>COM</td>
<td>MCQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>25.4</td>
<td>Illustrate the use of question words learnt earlier. Identify and use question words: why and how often etc.</td>
<td>COM</td>
<td>MCQ</td>
</tr>
<tr>
<td></td>
<td>Articles</td>
<td>26</td>
<td></td>
<td></td>
<td>*Recall and apply the rules for the use of a and an. Choose between a or an before words that start with mute consonant letters. Identify and use the definite article the. Differentiate between use of definite and indefinite articles.</td>
<td>COM</td>
<td>MCQ</td>
</tr>
<tr>
<td></td>
<td>Verbs</td>
<td>27</td>
<td></td>
<td></td>
<td>*Recognize and use more action verbs from extended environment including other subjects in speech and writing.</td>
<td>COM</td>
<td>MCQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27.1</td>
<td>*Demonstrate the use of be, do and have as main or helping verbs in sentences.</td>
<td>COM</td>
<td>MCQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27.2</td>
<td>*Illustrate use of can / cannot, may/ may not, and should / should not to express permission, prohibition, doubt, and obligation.</td>
<td>COM</td>
<td>MCQ/RRI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27.3</td>
<td>*Articulate and use forms of some simple regular verbs including be, do and have.</td>
<td>COM</td>
<td>MCQ/RRI</td>
</tr>
<tr>
<td></td>
<td>Tenses</td>
<td>28</td>
<td></td>
<td></td>
<td>* Illustrate the use of tenses (Simple present and continuous, simple past and continuous, and simple future tense) previously learnt in their speech and writing.</td>
<td>COM</td>
<td>MCQ/RRI</td>
</tr>
<tr>
<td></td>
<td>Words</td>
<td>29</td>
<td>Showing Position, Time and Movement (Prepositions)</td>
<td></td>
<td>*Demonstrate use of words showing position, time and movement and direction.</td>
<td>KNW</td>
<td>MCQ/RRI</td>
</tr>
<tr>
<td></td>
<td>Capitalization</td>
<td>30</td>
<td></td>
<td></td>
<td>*Use capitalization according to rules learnt earlier.</td>
<td>KNW</td>
<td>MCQ/RRI</td>
</tr>
<tr>
<td></td>
<td>Punctuation</td>
<td>30.1</td>
<td></td>
<td></td>
<td>*Recall the rules of punctuation learnt earlier.</td>
<td>KNW</td>
<td>MCQ/RRI</td>
</tr>
<tr>
<td></td>
<td>Joining Words (Conjunctions)</td>
<td></td>
<td></td>
<td></td>
<td>*Demonstrate use of joining words learnt earlier. *Recognize function of more joining words such as for example, for instance, to introduce an example etc.</td>
<td>KNW</td>
<td>MCQ/RRI</td>
</tr>
</tbody>
</table>
Find the following:

<table>
<thead>
<tr>
<th>Sr No</th>
<th>Content</th>
<th>Sub Content</th>
<th>SLOs No</th>
<th>SLO</th>
<th>Cognitive level</th>
<th>Que type</th>
<th>No of Que in each strand and its detail</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sentence Structure</td>
<td>32</td>
<td>*Recognize and use simple SVO pattern sentences with direct and indirect objects.</td>
<td>KNW</td>
<td>MCQ/RRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Types of Sentences</td>
<td>33</td>
<td>*Identify and make sentences to show instructions, commands and strong feelings.</td>
<td>COM</td>
<td>MCQ/RRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Types of Sentences</td>
<td>33.1</td>
<td>Recognize function of what forms used in questions.</td>
<td>KNW</td>
<td>MCQ/RRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Types of Sentences</td>
<td>33.2</td>
<td>Respond to, and ask wh questions.</td>
<td>COM</td>
<td>MCQ/RRI</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adjectives</td>
<td>34</td>
<td>*Classify adjectives of quantity, quality, size, shape colour and orgin. * Articulate, identify and use degrees of regular and irregular adjectives.</td>
<td>COM</td>
<td>MCQ/RRI</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Development of Table of Specification**

TOS (Table of Specification) is a practical word given to the plan for scripting items for a test. A table of specification should imitate what has been imparted in the instructional order. In other terms, the testing mode is a reflection of the instructional mode. The TOS should inculcate and reflect content and process, as the instructional mode of TOS has two basic dimensions - content matter and intellectual process. Bloom’s Taxonomy categories can be used to express the process, while developing TOS we advance with a framework which does not only reflect what has been imparted to the students, but also what intellectual level students possess.

- **Select a topic**
  What to be taught? Must be sorted at first to be specific and definite to proceed

- **Identify the sub-topics to be tested**
  What are the sub heads you will be covering in that topic to unfold the selected topic.
• **Identify the levels of the taxonomy that are most appropriate to assess the content**
  Based on the chosen topic and sub topics, the level or tier of taxonomy must be identified on which teacher or trainer has to work.

• **Review the types of skills that can be tested under each level**
  Keeping in view the requirements of the topic teacher must select the level of taxonomy under whose umbrella he/she is going to test the students.

• **Consider the number of items that will be used on the entire test and the time available.**
  Enlist the total number of test items and their required time to fill them.

• **Based on the time spent on/importance of each sub-topic, decide on the number of items that will be used for each sub-topic.**
  Based on the requirement of time for each sub topic, finalize the number of items to be used for judging student.

• **Distribute the questions across the cognitive levels**
  Divide the items on the basis of different level of cognitions.

iii. **Usage of Table of Specification**
  After the development of Table of Specification, the next big step is in knowing its elements and using it properly. The table of specification enables a teacher to test the various concepts and gauge them over Knowledge, Skill and Attitude. The table stipulates the suitable number of test items in the suitable content categories. However it further helps the teacher to elaborate the items in more descriptive way to be able to be specific in conveying the message and also defining the level of intellectual functioning. This table may also assist the instructor in determining the format and design of test items which will help him/her in identification of the weightage of the test items individually based on the respective modules.

iv. **Preparing the Test Blue-Print or Table of Specifications**
  Table of specifications is a two-way chart which relates the instructional objectives to the course content with emphasis on each learning outcome. It is a horizontal dimension labeled across the top with different behaviors that the teacher wants the students to display. Every technical design or development in most technical profession requires a blue print. According to Agbaegbu, Ezeudu and Agwagah (2004), it is a process which involves assigning weightages to topics in the content based on the instructional objectives specified by the syllabus or curriculum. The dimension consists of the course content in rows and the instructional objectives in the columns, sampling the desired cognitive to be tested.
The cognitive levels cover knowledge, comprehension, application, analysis, synthesis and evaluation. Knowledge and comprehension are simple processes while application analysis, synthesis and evaluation are complex processes (Inko-Tariah, 2004).

Table 1.
Table of specification or test blue-print for achievement test in Test and Measurement in school

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Definition of basic concepts of test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>concepts of test 30%</td>
<td>20%</td>
<td>25%</td>
<td>20%</td>
<td>15%</td>
<td>10%</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>Classification of test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20%</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Types of test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Essay and objective</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tests 25%</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Uses of test 15%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>Purpose of test 10%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Total No. of items</td>
<td>13</td>
<td>16</td>
<td>13</td>
<td>9</td>
<td>6</td>
<td>6</td>
<td>63</td>
</tr>
</tbody>
</table>

Another period of 20 percent of the time was for classification of test, 25 percent was spent on types of test, 15 percent was spent on uses of test, while 10 percent was on purpose or function of tests. Each of the behavioral instructional objectives was developed with knowledge 20 percent, comprehension 25 percent, application 15 percent, analysis 10 percent, synthesis 10 percent and evaluation 10 percent. These assignments were done on the basis of the emphasis placed on each instructional objective. It should be noted that the weight ages assigned to each cell is in essence the number of tests items which would be selected to determine the mastery of the topic. This is also dependent on the extent to which the instructional objectives have been realized. The preparation of the test blue-print is much more essential because it ensures that the students are assessed on all the relevant aspects of the subject matter. Again, the tests blue-print assists in eliminating the tendency to overload the test with items that cover very insignificant parts of the instructional objective, thereby and maintaining the content validity of the test.

v. Table of Specifications Using Bloom’s Revised Taxonomy

Instructions: Look at the first question on your test. First determine which of your objectives it is correlated with. Second, look at the verb in the question to determine which level of the taxonomy it is correlated with. Put the number of the question in the box that corresponds to the correct objective and level. Complete this process for each question on your test.
<table>
<thead>
<tr>
<th>Level of Cognitive Domain</th>
<th>Action Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remembering</td>
<td>Recall, identify, recognize, acquire, distinguish, state, define, name, list, label, reproduce, order, indicate, record, relate, repeat, select, tell, describe, match, locate, report, choose, cite, define, outline, complete, draw, find, give, isolate, pick, put, show</td>
</tr>
<tr>
<td>Recalling items of Information</td>
<td>Translate, extrapolate, convert, interpret, abstract, transform, select, indicate, illustrate, represent, formulate, explain (who/what/when/where/that/how), classify, describe, discuss, express, identify, locate, paraphrase, recognize, report, restate, review, summarize, find, relate, define, clarify, diagram, outline, compare, contrast, derive, arrange, estimate, extend, generalize, give examples, ask, distinguish</td>
</tr>
<tr>
<td>Level of Cognitive Domain</td>
<td>Action Verbs</td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Applying</td>
<td>Apply, sequence, carry out, solve, prepare, operate, generalize, plan, repair, explain, predict, instruct, compute, use, perform, implement, employ, solve, construct, demonstrate, give examples, illustrate, interpret, investigate, practice, measure, operate, adjust, show, report, paint, draw, collect, dramatize, classify, order, change, write, manipulate, modify, organize, produce, schedule, translate, complete, examine, advocate, persuade, resolve</td>
</tr>
<tr>
<td>Using information in new situations</td>
<td>Analyze, estimate, detect, classify, discover, discriminate, explore, distinguish, catalogue, investigate, break down, order, determine, differentiate, dissect, examine, interpret, calculate, categorize, debate, diagram, experiment, question, solve, test, dissect, deconstruct, focus, find coherence, survey, compare, contrast, classify, investigate, outline, separate, structure, categorize, determine evidence/premises and conclusions, appraise, criticize, debate, illustrate, infer, inspect, inventory, select, deduce, induce, argue, balance, moderate, identify, explain (how/why), challenge, question</td>
</tr>
<tr>
<td>Analyzing</td>
<td>Write, plan, integrate, formulate, propose, specify, produce, organize, theorize, design, build, systematize, combine, summarize, restate, discuss, derive, relate, generalize, conclude, produce, arrange, assemble, collect, compose, construct, create, perform, prepare, propose, strategize, compare, contrast, hypothesize, invent, discover, present, write, deduce, induce, bring together, pretend, predict, strategize, modify, improve, set up, adapt, solve, categorize, devise, explain (why), generate, manage, rearrange, reconstruct, relate, reorganize, revise, argue, extend, project, advocate, persuade, resolve</td>
</tr>
<tr>
<td>Distilling and/or organizing information into its components; solving problems</td>
<td>Evaluate, argue, verify, assess, test, judge, rank, measure, appraise, select, check, justify, determine, support, defend, criticize, critique, weigh, assess, choose, compare, contrast, decide, estimate, grade, rate, revise, score, coordinate, select, choose, debate, deduce, induce, recommend, monitor, compare, contrast, conclude, discriminate, explain (why), interpret, relate, summarize, challenge, question, advocate, persuade</td>
</tr>
</tbody>
</table>
SOLO Action Verbs

Common Verbs for Uni-structural Learning Outcomes
Paraphrase, define, identify, count, name, recite, follow simple instructions, calculate, reproduce, arrange, recognize, find, note, seek, sketch, pick

Common Verbs for Multi-structural Learning Outcomes
Combine, classify, structure, describe, enumerate, list, do algorithm, apply method, account for, execute, formulate, solve, conduct, prove, complete, illustrate, express, characterize

Common Verbs for Relational Learning Outcomes
Analyze, compare, contrast, integrate, relate, explain causes, apply theory (to its domain), argue, implement, plan, summarize, construct, design, interpret (some senses), structure, conclude, substantiate, exemplify, derive, adapt

Common Verbs for Extended Abstract Learning Outcomes
Theorize, generalize, hypothesize, predict, judge, transfer theory (to new domain), assess, evaluate, interpret (some senses), critically reflect, predict, criticize, reason

Test Grid

<table>
<thead>
<tr>
<th>Sr.</th>
<th>Student Learning Outcomes</th>
<th>SOLO</th>
<th>Pre</th>
<th>Uni/</th>
<th>Multi</th>
<th>Rela/</th>
<th>Ext</th>
<th>MCQ /CRQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rm</td>
<td>Un</td>
<td>App</td>
<td>Ana</td>
<td>Eva</td>
<td>Crt</td>
</tr>
<tr>
<td>1</td>
<td>Determine the perimeter and area of a square.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Match the numbers 0 – 9 with objects.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Identify the concepts related to Pakistan ideology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>State the key features of first SIMLA conference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Explain the term algebra</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Define isotopes.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Distinguish between shells and sub shells.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Narrate the early problems with special emphasize on economics and refugees</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Summarize the properties of liquids</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Convert the Molarity of a solution in g/dm³</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Use the rule that “like dissolves like”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Prepare solutions of different strength.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Sketch an electrolytic cell, label the cathode and anode.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Classify substances as Lewis acids or bases</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sr.</td>
<td>Student Learning Outcomes</td>
<td>SOLO</td>
<td>Pre</td>
<td>Uni/</td>
<td>Multi</td>
<td>Rela/</td>
<td>Ext</td>
<td>MCQ/CRQ</td>
</tr>
<tr>
<td>-----</td>
<td>---------------------------</td>
<td>------</td>
<td>-----</td>
<td>------</td>
<td>-------</td>
<td>-------</td>
<td>-----</td>
<td>---------</td>
</tr>
<tr>
<td>15.</td>
<td>list the seven units of System International</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>measure the thickness of a metal strip</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>plot and interpret distance-time graph</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>demonstrate various types of motion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>interpret graph from newspaper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>Analyze and compare formal and informal emails to note difference conventions, vocabulary, style and tone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>prove that potential energy Ep = mgh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>Outline gaseous exchange in plants,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>Analyze the 1970 elections</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>Write and revise letters for various purposes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25.</td>
<td>Identify the key aspects of 1973 constitution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26.</td>
<td>Estimate of different cells under the microscope</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27.</td>
<td>Interact with text and use reading strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28.</td>
<td>Point out common nouns in a English/Urdu text passage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SOLO**

[Pre=pre-structural, Uni=Uni-structural, Multi=Multi-structural, Rela=Relational, Exted=Extended Abstract]

**BLOOM**

[Rm= Remembering, Un= Understanding, Appli=Applying, Ana=Analyzing, Eva=Evaluating, Syn=Synthesize, Crt=Creat]

### 4.6 CONSTRUCTING A TABLE OF SPECIFICATION BASED ON SOLO TAXONOMY

The SOLO (Structure of the Observed Learning Outcome) model provides a systematic way of describing how a student’s performance grows in complexity when mastering tasks. It can be used in two ways:

i. To assess the learning outcomes attained by each student; and
ii. To set learning outcomes appropriate to where a student should be at a particular stage of their program.

It is relatively easy to apply the SOLO model to categorize students’ responses. The levels of the SOLO model can be used to construct a series of four levels relating to a single text, with each item measuring one level of the SOLO taxonomy. Consider the following super item that has four items, one at each level of the taxonomy:

**Concert Hall:**
The first row of the Mega Mall concert hall has 10 seats. Each row thereafter has 2 more seats than the row in front of it.

<table>
<thead>
<tr>
<th>Levels</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unistructural:</strong></td>
<td>The item requires that the student identify the next term in the sequence by referring directly to the information given.</td>
</tr>
<tr>
<td>How many seats are in row 2 in the Mega Mall concert hall?</td>
<td></td>
</tr>
<tr>
<td><strong>Multistructural:</strong></td>
<td>The item requires that the given information be handled serially. The student identifies the recursive relationship between the terms to solve the specific cases.</td>
</tr>
<tr>
<td>How many seats are in row 13 and 18 in the concert hall?</td>
<td></td>
</tr>
<tr>
<td><strong>Relational:</strong></td>
<td>The item requires that the student integrate all of the given information to make a generalization by forming an algebraic expression and linear equation. The student has to apply the rule to solve the related situation.</td>
</tr>
<tr>
<td>a. How many seats are in rows in the concert hall?</td>
<td></td>
</tr>
<tr>
<td>b. Write a linear equation to find the number of seats for any row. Let’s represents the number of seats and represents the number of rows.</td>
<td></td>
</tr>
<tr>
<td>c. If the last row has 406 seats, try to use the linear equation to find the number of rows in the hall.</td>
<td></td>
</tr>
<tr>
<td><strong>Extended Abstract:</strong></td>
<td>The item requires that the student extend his/her understanding of linear relationships by evaluating the relevance and applicability of the linear equation in the related situation and by forming an appropriate alternative solution for it.</td>
</tr>
<tr>
<td>The manager planned to prepare 1000–15000 seats in 100 rows for a musical concert. Will he make it? If yes, explain your answer. If no, try to suggest a new linear equation in order to help the manager.</td>
<td></td>
</tr>
</tbody>
</table>
4.7 WRITING TEST ITEMS (SUPPLY AND SELECTION TYPES) BASED ON TABLE OF SPECIFICATION

Selection-type test items can be designed to measure a variety of learning outcomes from simple to complex. It tends to be favored in achievement testing by

- Greater control of type of response students can make
- Broader sampling of achievement
- Quicker and more objective scoring

Despite these advantages, Supply-type test items can play an important role in measuring achievement. Supply-type test items require students to produce the answer, which may be single word or several page responses.

Length of response ranges along a continuum, but usually divided into: short answer, restricted-response essay, and extended-response essay.

<table>
<thead>
<tr>
<th>Supply-Type Item Continuum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highly Controlled Response</td>
</tr>
<tr>
<td>Outcomes Measured</td>
</tr>
<tr>
<td>Knowledge</td>
</tr>
<tr>
<td>Computational Skill</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

4.7.1 Fill in the Blanks

Completion items are especially useful in assessing mastery of factual information when a specific word or phrase is important to know. They preclude the kind of guessing that is possible on limited-choice items since they require a definite response rather than simple recognition of the correct answer. Because only a short answer is required, their use on a test can enable a wide sampling of content. A completion item requires the student to answer a question or to finish an incomplete statement by filling in a blank with the correct word or phrase. For example, According to Freud, personality is made up of three major systems, the_______, the ________ and the ________.

Completion items tend to test only rote, repetitive responses and may encourage a fragmented study style since memorization of bits and pieces will result in higher
scores. They are more difficult to score than forced-choice items and scoring often must be done by the test writer since more than one answer may have to be considered correct.

A distinction should be made between completions—often referred to as fill-in-the-blank—and short answer questions. With completion questions the response is usually one or two words that fit on a line provided by the tester. Short answer questions may require one sentence or even a paragraph to fully answer the question.

**Suggestions for writing completion items**

a) Omit only significant words from the statement.
b) Do not omit so many words from the statement that the intended meaning is lost.
c) Avoid obvious clues to the correct response.
d) Be sure there is only one correct response.
e) Avoid grammatical clues to the correct response.
f) If possible, put the blank at the end of a statement rather than at the beginning.

**4.7.2 Short Answer Questions and Marking Guide**

SAQs require the examinees to supply appropriate words, numbers or symbols to answer a question or complete a statement e.g.

- What are the incorrect responses in a multiple-choice item called? *(distractors)*
- The incorrect responses in a multiple-choice item are called *(distractors)*

SAQs can include computational problems and any simple item form that requires the answer rather than selecting it and primarily used to measure simple recall of knowledge. Short-answer appears easy to write but may be difficult to construct so only one answer is correct because several other defendable answers could be given, spelling may be a problem:

- does word have to be spelled correctly?
- if so may penalize poor speller who knows answer
- if not may have difficulty determining if misspelled word is the correct response (requires subjectivity)

Due to these difficulties short-answer items should be reserved to recall specific information where the answer is obvious.

i. **Short-Answer Items Strengths**
   a) It is easy to write test items
b) Guessing is less likely than in selection-type test items
c) This test item type is well-suited to computational problems and other learning outcomes where supplying the answer is important
d) A broad range of knowledge outcomes can be measured

ii. Short-Answer Items Limitations

a) It is difficult to phrase statements so that only one answer is correct
b) Scoring is contaminated by spelling ability except when responses are verbal
c) Scoring is tedious and time-consuming
d) This item is not very adaptable to measuring complex learning outcomes

iii. Rules for Writing Short-Answer Items

a) State the item so that only a single, brief answer is possible
   • Requires skill in phrasing and precise terms
   • What appears as simple and clear questions can often be answered in different ways
   • Review the item with this in mind
b) Start with a direct question and switch to an incomplete statement only when greater conciseness is possible by doing so
   • The use of direct questions increases likelihood that problem will be stated and that only one answer will be appropriate
   • Incomplete sentences tend to be less ambiguous when based on problems first stated as sentences

Example
What is another name for true-false items? (alternative-response items)
True-false items are also called (alternative-response items)

c) It is best to leave only one blank, and it should relate to the main point of the statement
   • Leaving several blanks is often confusing and the answer to one blank may depend on the answer in another

Examples (Poor and Better)
In terms of response, the (matching item) is the most like (multiple-choice items)
In terms of type of responses, which item is the most like matching item? (multiple-choice)
   • It is important to avoid asking students to respond to unimportant aspects of a statement

d) Place the blanks at the end of the statement
   • This allows student to read the complete problem before coming to the blank to be filled (avoids having to re-read the sentence)
   • Easier to construct incomplete statements
Examples (Poor and Better)

- The supply-type item used to measure the ability to organize and integrate material is called an *(essay item)*
- Supply-type items used to measure the ability to organize and integrate material are called *(essay items)*

e) *Avoid extraneous clues to the answer (i.e., extra unnecessary information)*

- A Common clue is length of the blank (e.g., long blank for long answer, short blank for short answer)
  - All blanks should be uniform in length
- Another common clue is use of the indefinite article “a” or “an” just before the blank

Examples (Poor and Better)

- The supply-type item used to measure the ability to organize and integrate material is called an *(essay item)*
- Supply-type items used to measure the ability to organize and integrate material are called *(essay items)*
- Poor version (the article “an”) eliminates the only other supply-type item choice “short-answer”
  - Solution - eliminate the article “an” or include both articles a(an)
  - Use plural verb as in Better example

f) *The numerical answers indicate the degree of precision expected and the units in which they are expressed*

- Example “nearest the whole number”
- Specify the degree of precision prevents the student from spending too much time on the problem
- Provides more uniform responses

iv. Checklist for Evaluating Short-Answer Items

1. Is this type of item appropriate for measuring the intended learning outcome?
2. Does the item task match the learning task to be measured?
3. Does the item call for a single, brief answer?
4. Has the item been written as a direct question or a well-stated incomplete sentence?
5. Does the desired response relate to the main point of the item?
6. Is the blank placed at the end of the statement?
7. Have clues to the answer been avoided (e.g., “a” or “an”, length of blank)
8. Are the units and degree of precision indicated for numerical answers?
4.7.3 Essay Type Questions and Rubric for Marking

- Provide greater freedom of response for students as to:
- Choice on how to approach problem
- What factual information to use?
- The degree of emphasis to place on each aspect of the response
- Essay questions especially useful for measuring the ability to organize, integrate and express ideas
- Areas where selection-type items fall are inadequate
- Essays and selection-type items complement each other

i. Summary Comparison of Selection-Type Items and Essay

<table>
<thead>
<tr>
<th>Selection-Type Items</th>
<th>Essay Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Outcomes Measured</td>
<td>Good for measuring the recall of knowledge understanding and application levels of learning, inadequate for organizing and expressing ideas</td>
</tr>
<tr>
<td>Sampling of Content</td>
<td>The use of a large number of items results in broad coverage, which makes representation sampling of content feasible</td>
</tr>
<tr>
<td>Preparation of Items</td>
<td>Preparation of good items is difficult and time consuming</td>
</tr>
<tr>
<td>Scoring</td>
<td>Objective, simple and highly reliable</td>
</tr>
<tr>
<td>Factors Distorting Scores</td>
<td>Reading ability and guessing</td>
</tr>
<tr>
<td>Probable Effect on Learning</td>
<td>Encourages students to remember, interpret and use the ideas of others</td>
</tr>
<tr>
<td></td>
<td>Ineffective for measuring the recall of knowledge; best for the ability to organize, integrate and express ideas</td>
</tr>
<tr>
<td></td>
<td>The use of a small number of items limits coverage, which makes representative sampling of content infeasible</td>
</tr>
<tr>
<td></td>
<td>Preparation of good items is difficult but easier than selection-type items</td>
</tr>
<tr>
<td></td>
<td>Subjective, difficult and less reliable</td>
</tr>
<tr>
<td></td>
<td>Writing ability and bluffing</td>
</tr>
<tr>
<td></td>
<td>Encourages students to organize, integrate and express their own ideas</td>
</tr>
</tbody>
</table>

ii. Types of Essay Questions

b) Restricted Response Questions
- Places strict limits on the answers to be given as defined
• Boundaries of subject matter narrowly defined by the problem and specific form of the answer
  ○ Indicated by words as “list,” “define,” and “give reasons”

Example
Describe the relative merits of selection-type items and essay questions for measuring learning outcomes at the understanding level. Confine your answer to one page.

Mr. Rogers a ninth-grade science teacher, wants to measure his students’ “ability to interpret scientific data” with a paper-and-pencil test
1. Describe the steps that Mr. Rogers would follow,
2. Give reasons to justify each step

c) Extended Response
• Provide students almost unlimited freedom to determine the form and scope of their response
  ○ Some practical limits may be imposed such as time and page limits
• Students given sufficient freedom to demonstrate skills of evaluating and creating

Example (Evaluation Outcome):
(The student is given a complete achievement test that includes errors or flaws in the directions, in the test items and in the arrangement of the items)
Write a critical evaluation of this test using as evaluative criteria the rules and standards for test construction described in your textbook. Include a detailed analysis of the test’s strengths and weaknesses and an evaluation of its overall quality and probable effectiveness.

d) Essay Questions Strengths
i. The highest level learning outcomes (analyzing, evaluating, creating) can be measured
ii. Preparation time is less than for selection-type items
iii. The integration and application of ideas is emphasized

e) Essay Questions Limitations
i. There is an inadequate sampling of achievement in time needed for answering each question
ii. It is difficult to relate to intended learning outcomes because of freedom to select, organize and express ideas
iii. Scores are raised by writing skill and bluffing and lowered by poor handwriting, misspelling and grammatical errors
iv. Scoring is time consuming, subjective and tends to be unreliable
f) **Rules for Writing Essay Questions**
Use essay questions to measure complex learning outcomes only
Relate the question as directly as possible to the learning outcome being measured
i. Formulate questions that present a clear task to be performed
ii. Do not permit a choice of questions unless the learning outcomes require it
iii. Provide ample time for answering and suggest a time limit on each question

g) **Rules for Scoring Essay Answers**
i. Evaluate answers to essay questions in terms of the learning outcomes being measured
ii. Score restricted-response answers by the point method, using a model answer as a guide
iii. Score extended-response answers by the rating method, using defined criteria as a guide
iv. Evaluate all the students’ answers to one question before proceeding to the next question
v. Evaluate answers to essay questions without knowing the identity of the writer
vi. Whenever possible, have two or more persons grade each answer

h) **Student Bluffing and Scoring Essays**
Students can obtain higher scores on essay questions by clever bluffing. Although this requires skill in writing and some knowledge of the topic, credit should not be given unless the question is specifically answered. Some common types of bluffing are listed below
i. Student repeats the question in statement form (slightly paraphrased) and tells how important the topic is (e.g., the role of assessment in teaching is extremely important. It is hard to imagine effective instruction without it, etc.)
ii. Student writes on a well-known topic and fits it to the question (e.g., a student who knows testing, well but knows little about performance assessment and is asked to compare testing and performance assessment might describe testing in considerable detain and frequently state that performance assessment is much superior to evaluating the type of leaning assessment by the tests)
iii. Student liberally sprinkles the answer with basic concepts regardless of whether they are understood (e.g., when asked to write about any assessment technique, the importance of “validity” and “reliability” is mentioned frequently)
iv. Student include the teacher’s basic beliefs whenever possible (e.g., “The intended learning outcomes must be stated in performance terms before this type of test is constructed or selected”)

4.8 SELF ASSESSMENT QUESTIONS

1. Highlight the role of assessment in teaching and learning process.
2. What factors should be considered in deciding whether subjective or objective type questions should be included in classroom tests?
3. Describe the various types of reporting test score by giving examples from our country context.
4. How do you define an objective and an outcome? Differentiate between objectives and outcomes with the help of examples.
5. Enlist the different types of techniques and their role in education system.
BIBLIOGRAPHY


Fleming, N.D. and Mills, C. (1992), Not Another Inventory, Rather a Catalyst for Reflection, To Improve the Academy,11(3),137-147.


DEVELOPMENT OF SELECTION TYPE TEST ITEMS

Written By: Safia Janjua
Reviewed By: Dr. Muhammad Zafar Iqbal
# CONTENTS

Introduction ......................................................................................................................... 99  
Objectives .......................................................................................................................... 99  

5.1 Need and Scope of Selection Type Test Item ......................................................... 100  
5.2 Multiple Choice Items (MCQs) ........................................................................... 100  
5.3 Matching Items ....................................................................................................... 102  
5.4 True False Test Items ............................................................................................. 103  
5.5 Development of Item Review Criteria and Review Of The Items .................. 105  
5.6 Self Assessment Questions ...................................................................................... 109  
Bibliography ..................................................................................................................... 110
INTRODUCTION

Classroom test and assessments play a central role in the evaluation of student learning. The main goal of classroom testing and assessment is to obtain valid, reliable and useful information concerning student achievement. Writing items requires a decision about the nature of the item or question to which we ask students to respond. A test item is a specific task, test takers are asked to perform. These items are required the students to select a correct or best answer. These items can be scored more objectively. These items allow adequate content sampling because they require less time on the part of examinee rather than subjective tests.

In selection type test items, the respondents have to select the responses from provided alternatives. These test items consist of declarative statements in which the students have to point out the true or false, right or wrong, or correct or incorrect statements according to the directions of the test developer. Therefore, this type of test items is comparatively easy to answer than the supply type test item.

OBJECTIVES

After studying this unit, prospective teachers will be able to:

- understand, comprehend and apply the principles of selection type test items.
- identify and evaluate these different methods of selection type test items.
- construct different test items while keeping in mind their items criteria.
5.1 NEED AND SCOPE OF SELECTION TYPE TEST ITEMS

In selection type test items several possible answers/alternatives for each question are already given and the students are only required to select the correct or best answer in them instead of recalling facts or information from their own memories. This type of test items consist of declarative statements in which the students have to point out the true or false, right or wrong or correct or incorrect statements according to the directions of the test developer. Selection type test items are generally consist of the following types.

- Writing MCQs
- Matching Column
- True/False

5.2 MULTIPLE CHOICE ITEMS (MCQS)

A multiple-choice item asks a question or establishes the situation for a response. This type of item usually includes three or four response choices, or options from which the correct answer is selected. A multiple-choice item is characterized by the following components:

- The stimulus presents the contextual information relevant to the item.
- The stem presents the question or prompt the student must answer.
- The options refer to the entire set of labeled response choices presented under the stem.
- The key is the correct response option.
- The distracters are the incorrect response options.
For Example:
Which planet is closest to the sun?
A. Venus
B. Earth
C. Jupiter
D. Mars

Characteristics of Multiple-Choice Items
- Versatility in measuring all levels of cognitive ability.
- Highly reliable test scores.
- Scoring efficiency and accuracy.
- Objective measurement of student achievement or ability.
- A wide sampling of contents or objectives.
- A reduced guessing factor when compared to true-false items.
- Incorrect response alternatives can provide diagnostic feedback.

Limitations in Using Multiple-Choice Items
- Are difficult and time consuming to construct.
- Lead an instructor to favor simple recall of facts.
- Place a high degree of dependence on the student's reading ability and instructor's writing ability.
- Frequently difficult to find plausible distracters.
- Scores can be influenced by reading ability.
- Often focus on testing factual information and fails to test higher levels of cognitive thinking.

Activity 1
Develop a test based on multiple Choice items from a specific chapter and evaluate it with different characteristics of multiple choice items.

Suggestions for writing Multiple Choice Test Items
1. The stem of the item should clearly formulate a problem.
2. Be sure that there is one and only one correct or clearly best answer.
3. Be sure wrong answer is plausible.
4. Include from three to five options to optimize testing for knowledge rather than encouraging guessing.
5. To increase the difficulty, increase the similarity of content among the options.
6. Use the caption “none of the above” sparingly and only when the keyed answer can be classified unequivocally as right or wrong.
7. Avoid using “all of the above”. In this type, a question had to be answered by the respondent by selecting the correct response out of the alternative responses.

8. Avoid using negative stems—those containing words such as NOT, LEAST, WORST, EXCEPT, etc.

5.3 MATCHING COLUMN TEST ITEMS

Matching test items consist of
- A column of premises
- A column of responses
- Directions of matching the two.

The matching test item format provides a way for learners to connect a word, sentence or phrase in one column to a corresponding word, sentence or phrase in a second column. The items in the first column are called premises and the answers in the second column are the responses. Each item in one list is paired with at least one item in the other list. Matching may be considered to be a variant of multiple-choice in which more than one choice is correct. Matching test items are usually used for measuring the ability of students to identify a connection/relationship between two things. This relationship may exist in the form of time, date, place, events, class, physical characteristics etc.

For Example:

<table>
<thead>
<tr>
<th>Premises</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column A</td>
<td>Column B</td>
</tr>
<tr>
<td>1. Person who performs mysterious tasks no one understands</td>
<td>A. Facilitator</td>
</tr>
<tr>
<td>2. Person who provides schooling for children</td>
<td>B. Trainer</td>
</tr>
<tr>
<td>3. Person who enables a group to find solutions</td>
<td>C. Instructional Designer</td>
</tr>
<tr>
<td>4. Person who instructs adults in a classroom.</td>
<td>D. Meeting Organizer</td>
</tr>
<tr>
<td></td>
<td>E. Teacher</td>
</tr>
</tbody>
</table>

Characteristics of Matching Items
- Require short periods of reading and response time, allowing you to cover more content.
- Provide objective measurement of student achievement or ability.

102
• Provide highly reliable test scores.
• Provide scoring efficiency and accuracy.
• Allow for great flexibility and accuracy in counting the learners’ scores.
• Give an objective assessment of the learners’ knowledge.
• At their most useful when used in areas mostly dealing with facts.
• Least chance of guessing the correct answer compared to other question types.

Limitations of Matching Items
• Have difficulty measuring learning objectives requiring more than simple recall of information.
• Are difficult to construct due to the problem of selecting a common set of stimuli and responses.
• Students can use rote memorization to answer these exercises, especially as typical matching problems involve assessment of rote associations such as names, dates, etc.
• Can be difficult to develop homogeneous premises and responses.

<table>
<thead>
<tr>
<th>Activity 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop any matching item test from any specific topic while keeping in mind its strength and limitations.</td>
</tr>
</tbody>
</table>

Suggestions for writing Matching Type Test Items
1. Use only homogeneous items in the list of premises (items for which a match is sought) and responses (items from which a selection is made).
2. The number of responses should be more than the number of premises.
3. Place the words in alphabetical order and numbers in sequence in the list of premises and responses.
4. Direct the students that they can use an item in the list of responses once, more than once or not at all.
5. Tell the students about the basis upon which the match is sought.
6. Place responses to the right of premises.
7. Each choice should be about the same length.
8. State the stem of the item in positive form, whenever possible.

5.4 TRUE-FALSE TEST ITEMS

True-false test items are typically used to measure the ability to identify whether or not the statements of facts are correct. The basic format is simply a declarative
statement that the student must judge as true or false. No modifications of this basic form in which the student must respond “YES” or “NO,” “AGREE” or “DISAGREE.” (Gronlund, 1998). The student must judge whether the sentence is a true or a false statement. A true-false item can be written in one of the three forms: simple, complex or compound. Answers can consist of only two choices (simple), more than two choices (complex), or two choices plus a conditional completion response (compound).

**For Example:**
Simple: The acquisition of morality is a developmental process. True/ False
Complex: The acquisition of morality is a developmental process. True/ False
opinion
Compound: The acquisition of morality is a developmental process. True/ False
If the statement is false, what makes it false?

**Advantages of True/False Items**
1. Scoring efficiency and accuracy
2. Versatility in measuring all levels of cognitive ability
3. Highly reliable test score
4. An objective measurement of student achievement or ability
5. They are relatively easy to write and can be answered quickly by students.
6. They provide the widest sampling of content per unit of time.

**Limitations of True/False Items**
1. The problem of guessing is the major weakness. Students have a fifty-percent chance of correctly answering an item without any knowledge of the content.
2. Items are often ambiguous because of the difficulty of writing statements that are unequivocally true or false.
3. Do not discriminate between of students of varying ability as well as other item types.
4. Can often include more irrelevant clues than do other item types.
5. Can often lead a teacher to favour testing of trivial knowledge.

<table>
<thead>
<tr>
<th>Activity 3:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop an objective test based on true/ false items for any class, administer it, and then analyze its item difficulty level.</td>
</tr>
</tbody>
</table>

104
Suggestions for writing True-False Test Items
1. Avoid writing a very long statement.
2. Construct statements that are definitely true of definitely false, without additional qualifications.
3. Avoid specific determiner such as “Never”, “Always”, “All”, “None”.
4. Keep true items must be the same with the number of false items.
5. Avoid using trick questions.
6. Avoid using double-negative statements. They take extra time to translate and are difficult to interpret.
7. Avoid grammatical clues that could lead to a correct answer such as the article (a, an, the).
8. Approximately half of the statements should be false (if the answer is false).

5.5 DEVELOPMENT OF ITEM REVIEW CRITERIA AND REVIEW OF THE ITEMS

How to write multiple choice items? Suggestion for writing multiple choice items:

1. Write multiple-choice items with three or four response options labeled A–C or D.
2. Make sure that one of the four response options, there is only one correct or best answer.
3. Make sure that the three or four response options are independent.
4. Make sure that the grammatical structure of all response options “fit” the stem.
5. Make sure all (or sets) of the response options are parallel in length, level of complexity and grammatical structure.
6. Do not use words or phrases in the stem that are repeated in one of the response options and, therefore, act as a clue to the correct response.
7. Do not use “none of these” and “all of these” as response options.
8. Arrange the response options in a logical order if this makes sense and saves the student time in reading the options (e.g., years in chronological order, numbers from least to greatest.

Plausibility of Distracters

Use plausible distracters (incorrect response options) that are based on likely student errors or misconceptions. This reduces the likelihood of students arriving at the correct response by eliminating other choices and, equally important, may allow identification of widespread student misunderstandings or tendencies that could lead to curricular or instructional improvements. If there are no plausible
errors or misconceptions, still make the options “reasonable.” For example, they
should be from the same area of content. However, avoid the use of “trick”
distracters.

For Constructing an Effective Stem

1. The Stem should be Meaningful by Itself
   A stem that presents a definite problem allows a focus on the learning
   outcome. A stem that does not present a clear problem, however, may test
   students’ ability to draw inferences from vague descriptions rather serving
   as a more direct test of students’ achievement of the learning outcome.

For Example: If Stem is not meaningful...

<table>
<thead>
<tr>
<th>Which of the following is not a true statement?</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Mitochondrial genomes are relatively constant in content (i.e., types of genes present)</td>
</tr>
<tr>
<td>B. Mitochondrial genomes are relatively constant in organization</td>
</tr>
<tr>
<td>C. Mitochondrial genomes are relatively constant in size</td>
</tr>
</tbody>
</table>

Better Stem...

<table>
<thead>
<tr>
<th>What characteristic is relatively constant in mitochondrial genomes across species?</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Content (i.e., types of genes)</td>
</tr>
<tr>
<td>B. Organization</td>
</tr>
<tr>
<td>C. Size</td>
</tr>
</tbody>
</table>

2. The Stem should be Negatively Stated only Required.

For Example:

<table>
<thead>
<tr>
<th>A water type extinguisher is suitable for putting out a fire caused by burning all of the following except</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Alcohol</td>
</tr>
<tr>
<td>B. Cotton</td>
</tr>
<tr>
<td>C. Paper</td>
</tr>
<tr>
<td>D. Wood</td>
</tr>
</tbody>
</table>

Source: Cheung, Derek and Bucat, Robert, How can we construct good multiple-choice items? Presented at the Science and Technology Education Conference, Hong Kong, June 20-21, 2002
Constructing an Effective Alternatives

3. All Alternatives should be Plausible.
Alternatives that are implausible don’t serve as functional distractors and thus should not be used. Common student errors provide the best source of distractors.

Who gathered the data that helped reveal the structure of DNA?

| A. | Francis Crick |
| B. | George Washington |
| C. | James Watson |
| D. | Rosalind Franklin |
| E. | Snoopy |

Note: B and E are not functional distracters so do not contribute the items.

- **How to Write Matching Column Items**
  - **Give Clear Instructions.**
    Let students know the basis on which items are to be matched, where to write answers and whether a response may be used more than once.
  - **Keep the two sets of items homogeneous.**
    For example, Column 1 may list events and Column 2 may list dates; do not combine events, dates and names in one column.
  - **Try to order the responses.**
    If you order the items in Column 2 alphabetically, chronologically, or conceptually, students will be able to read the series quickly and locate answers rapidly.
  - **Create more responses than premises.**
    In general, give students five to ten alternatives in Column 2. If you include distracters in Column 2, let students know that some of the entries in Column 2 do not apply.
  - **Be conscious of layout and format.**
    Always keep both columns on the same page so that students don't have to flip back and forth. Place answer blanks to the left of each entry in Column 1. Place Column 2 on the right-hand side of the page. Use capital letters for the responses (they are easier to discern than lowercase letters) and numbers for the premises (for later discussion).

**Best Practices for Constructing Matching Column Test**
  - Keep questions short and straightforward. Avoid unnecessary words.
  - Do not get carried away adding additional items. Having 10-12 items between both columns (5-6 “question - answer” pairs) is the sweet spot.
  - It is best to arrange the items in the left column according to some criterion (alphabetically, chronologically, etc).
• Make sure that no items in the right column can be matched to more than one item in the left one. However, having an item in the left column serves as the key for more than one item in the right column is all right.
• Avoid positioning matching test questions in such a way that the list is separated in two by a page break. Learners should not have to go back and forth trying to match questions on one page to answers on the other.
• When constructing answers, try to keep them interconnected by theme and the manner of presentation. You can find examples of correctly and incorrectly constructed matching questions below.

Good Example of Matching Column Question
Q: Column A contains a list of characteristics of questions. On the line to the left of each phrase, write the letter of the question type in Column B that best fits the phrase. Each response in Column B may be used once, more than once, or not at all.

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B) 1. Least useful for educational diagnosis.</td>
<td>A. Multiple-choice</td>
</tr>
<tr>
<td>(A) 2. Measures greatest variety of learning outcomes.</td>
<td>B. True-false</td>
</tr>
<tr>
<td>(C) 3. Most difficult to score objectively.</td>
<td>C. Short answer</td>
</tr>
<tr>
<td>(B) 4. Provides the highest score by just guessing.</td>
<td></td>
</tr>
</tbody>
</table>

Bad Example of a Matching Question
Q: Column A lists several phrases related to Pennsylvania. To the left of each phrase, write the letter of the response in column B that best fits the phrase.

<table>
<thead>
<tr>
<th>Column A</th>
<th>Column B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pennsylvania's official state flower</td>
<td>A. Ruffed grouse</td>
</tr>
<tr>
<td>2. Pennsylvania's official state bird</td>
<td>B. Pittsburgh</td>
</tr>
<tr>
<td>3. Major steel producing city</td>
<td>C. Mountain laurel</td>
</tr>
<tr>
<td>4. 2003 Penn State student population</td>
<td>D. Over 80,000</td>
</tr>
<tr>
<td>-</td>
<td>E. Erie</td>
</tr>
</tbody>
</table>

Main Errors
• Directions about how to handle choices in Column B are unclear.
• The responses are not homogeneous. That is, answers for 1, 2 and 4 are obvious, leaving only the two cities as choices for 3.

How to Write True/False
Suggestions for Writing True-False Test Items (PAYNE, 1984)
1. Based true-false items upon statements that are absolutely true or false, without qualifications or exceptions.
Poor: Near-sightedness is hereditary in origin.
Better: Geneticists and eye specialists believe that the predisposition to near-sightedness is hereditary.

2. **Express the items statement as simply as clearly as possible.**

   **Poor:** When you see a highway with a marker that reads: “Interstate 80,” you know that the construction and upkeep of that road is built and maintained by the local and national government.
   **Better:** The construction and the maintenance of the interstate highways are provided by both local and national government.

3. **Express a single idea in each test item.**

   **Poor:** Water will boil at a higher temperature if the atmospheric pressure on its surface is increased and more heat is applied to the container.
   **Better:** Water will boil at a higher temperature if the atmospheric pressure on its surface is increased. Or water will boil at a higher temperature if more heat is applied to the container.

4. **Include enough background information and qualifications so that the ability to respond correctly to the item does not depend on some special, uncommon knowledge.**

   **Poor:** The second principle of education is that the individual gathers knowledge.
   **Better:** According to John Dewey, the second principle of education is that the individual gathers knowledge.

5. **Avoid lifting statements from the text, lecture or other materials so that memory alone will not permit a correct answer.**

   **Poor:** For every action there is an opposite or equal reaction.
   **Better:** If you were to stand in a canoe and throw a life jacket forward to another canoe, chances are, your canoe will jerk backward.

6. **Avoid using negatively stated item statements.**

   **Poor:** The Supreme Court is not composed of nine justices.
   **Better:** The Supreme Court is composed of nine justices.

7. **Avoid trick questions.**

   **Poor:** “The Raven” was written by Edgar Allen Poe.
   **Better:** “The Raven” was written by Edgar Allan Poe.

**5.6 SELF ASSESSMENT QUESTIONS**

1. Compare the function of selection and supply type’s items.
2. State different types of reliability and explain each type with examples.
3. Define the term validity and elaborate its different types.
4. What strategies you adopt to plan an annual exam of your class?
5. What should be essentials of a good progress report? Discuss in detail with respect to public school system in Pakistan.
BIBLIOGRAPHY

Cheung, Derek and Bucat, Robert. How can We Construct Good Multiple-choice Items? Presented at the Science and Technology Education Conference, Hong Kong, June 20-21, 2002.

DEVELOPING ALTERNATE TECHNIQUES-I

Written by: Dr. Jahan Ara Shams
Reviewed By: Dr. Muhammad Zafar Iqbal
CONTENTS

Introduction ........................................................................................................................................113
Objectives .........................................................................................................................................113
6.1 Classroom Observations .............................................................................................................114
  6.1.1 Classroom observations: Purpose and Significance .................................................................114
  6.1.2 Planning and Preparing for Observation ..................................................................................115
  6.1.3 Typical Observation Formats ..................................................................................................117
  6.1.4 Making Decision on the Basis of Observation Results ............................................................123
6.2 Classroom Projects .....................................................................................................................124
  6.2.1 Definition of a Project ............................................................................................................124
  6.2.2 Tasks versus Tests ..................................................................................................................126
  6.2.3 Characteristics of a Project .....................................................................................................127
  6.2.4 Indicators of Successful Project ..............................................................................................128
6.3 Phases/steps of a project ..............................................................................................................129
  6.3.1 How Can Projects are implemented in a Classroom ...............................................................130
  6.3.2 Assessment of Projects and Use for Evaluation ......................................................................130
6.4 Methods of Calculate Reliability ..............................................................................................131
6.5 Self-Assessment Questions ........................................................................................................132
Bibliography .....................................................................................................................................133
INTRODUCTION

The process of education is not an open-ended process. Therefore, it requires the assessment of the students so that their progress can be analyzed from time to time during an educational program. This assessment has a number of facets which includes conventional examination techniques and testing procedures. With the passage of time the process of assessment has evolved and its requirements have changed with a great volume. This increasing requirement asks for developing alternate techniques for the assessment. These techniques should cover all the aspects of the educational process and they must be able to fulfill the criteria of a specific program.

OBJECTIVES

After studying this unit, prospective teachers will be able to;
- understand the process of classroom observation.
- create a proper plan to carry out the process of observation
- derive results from the process of observation
- explain classroom projects and implement in their classrooms
- describe different types and phases of projects
- assess the projects and use that information for project evaluation
- design different tasks and tests
- elucidate methods of calculating reliability
6.1 CLASSROOM OBSERVATION

Observation plays a vital role in assessment. It is an important component of learning process. Classroom observation can be termed as a continuous process of assessment. Teachers can assess different mind states of students which may be motivation, frustration or excitement. When the teacher gets direction from this observation he/she sets his plans accordingly. From the old days observation has been identified as an authentic source of information recording student’s progress. In order to get the proper benefits of the observation process, a proper system must be followed by the teacher. This systematic process requires an adequate preparation and deep thought process. By this it is not implied that the each and every aspect of observation process must be taken into the account but our approach must be well thought and it should not be an accidental approach. It is necessary to know the outcomes of learning process in advance and how the required data can be collected. Observation allows teachers to focus on how students learn and respond in particular setting. During teaching many things occur besides learning which could not be neglected. Observation helps to improve performance and provides positive outcomes. We can categorize classroom observation in two categories i.e. incidental observation and planned observation.

Incidental observation occurs during the ongoing (deliberate) activities of teaching and learning and the interactions between teacher and students. We can say that in the classroom context an unplanned situation occurs where teacher observes some aspect of individual student learning. The use of incidental observation for formal assessment and reporting depends upon the type of record, kept while this observation. Planned observation is carried out to get a specific result that can be obtained by providing a preplanned environment to the teacher and student. This observation can happen in a regular educational session or class or it may be carried out in a specially created environment.

6.1.1 Classroom Observations: Purpose and Significance

Through the process of observation teachers can collect the data regarding the progress of student in a number of authentic ways. Observation can lead a teacher towards a perfect relation between teacher and student. With the use of observation tool we can assess the current progress of student and speculate about his/her future progress. The most appropriate way to assess children, is to observe them while they are involved in activity. This process finely identifies the indicators of a child’s development which may be his/her likes and dislikes or strengths and weaknesses. Classroom observation makes assessment more comprehensive and to the point according to curriculum framework, pedagogical
planning and learning experiences. It connects the process of assessment to a specific context like gathering data from variety of situations. The assessment becomes more authentic with the help of classroom observation as it provides true and real picture to the teacher. The purposes of classroom observation are:

- To assess the child’s current state of learning.
- To assess the appropriateness of given set of instructions.
- To plan the lesson properly
- To provide external information (to parents, other staff members etc)
- To assess the level of communication among students and students or students and teachers.
- To assess behaviour, physical development, social interaction, conversation etc.
- To assess teacher’s performance, activities and interaction with students.
- To provide an adequate environment to aid the complete academic development of the students and to fulfill the requirement of each student.
- To take appropriate measures in case the student is not performing according to his/her age requirements.
- To maintain positive relation with students.
- To monitor curriculum, activities and outcomes and make improvements accordingly.

6.1.2 Planning and Preparing for Observation

Planning acts as a backbone of every process. Planning provides an authentic set of information that may serve as the best answer to the question of accountability. When a well planned observation is being carried out it provides an authentic source of recorded data. This recorded data can be used to prove the correctness of our derived results in front of any accountability forum, which may be parents, managers of educational facility. The teacher always comes to a judgment after the process of assessment. A well planned observation process provides very useful information that can be revisited or reanalyzed to verify the outcomes of assessment process. Observational evidences could be in different forms. A demonstrable outcome is the basic requirement of any form of evidence. This becomes more important when the observation is incidental and its results are recorded instead of recording the observation. A predefined set of objectives of learning outcomes can further help the teacher about the progress of the student through the process of his communication. The most important aspect of planning is learning about the structure, language and concept of framework of learning outcomes. A vital principal of planning observation is the true and actual understanding of the objective or purpose of the observation for which the process will be planned. The appropriateness of observation method depends upon the student’s age and I.Q. level. The process should be meaningful so that the desired
objective could be achieved. The ethical aspects must be kept in mind while planning. They key points to remember are: Why do there is a need of observation? What to observe? How to observe? How to assure the accuracy and authenticity of obtained information? The planning of observation goes through following steps:

- Decide the need and objectives
- Consider ethical issues
- Select appropriate method of observation
- Implementation of method
- Ways of deriving results
- Reflection on outcomes

While preparing for observation, the recording method should be kept in mind for both planned and incidental observation. A teacher must have clear understating of accurate record keeping method or type of observation record. It could be direct record, observation sheet or any other form. With the preparation of observation sheet we can have a systematic judgment which will add value to the teacher’s observation. Observation sheet must include a list of desired outcomes and the categories of students’ activities and performance. Learning outcomes on observation sheet must be elaborated clearly along with their judgment criteria. If learning outcomes are well defined with the help of their indicators and characteristics then the observation comes out in a well elaborated form. In order to conduct classroom observation a teacher must prepare following items to obtain maximum benefits. A teacher must have to prepare

**Artifacts**
A teacher’s observation has the primary objective of observing different activities and performances of the students. In some cases the production of an artifact can be termed as a result of that performance or activity and in other case there is no artifact produced and all the attention goes to the activity or process itself. The term artifact refers to anything that is constructed by a student for example a worksheet, an essay, a diagram, a painting, a composition or a webpage that can be said as a product. In teacher’s observation the subject is not the artifact but the subject is the way in which the artifact has been produced that is called the process.

**Direct record**
Direct record refers to any physical evidence which may be an audio or video recording or photography. In this case the event or activity must be like a speech, physical presentation, group activity or any practical task. This recording does not mean to record the whole event itself but in the best form it elaborates some
features or characteristics of that specific event that may be recalled in future. In this case we cannot cover the observation of whole event for example the feeling of that event or the passion of students and audience cannot be recorded. The technical aspects may also effect the quality of observation as we may face the problems with recording devices etc. Therefore it is implied that this type of record becomes the partial representation of the activity but it is thought better despite of having no record.

Written record
A written record can be produced in the form of an observation sheet which may be a log book or a diary. Observation sheets can be detailed or brief at some time. They may contain only a checklist of learning outcomes or a detailed on the spot observation of the event which contain minute details etc. In between these two types a combination of both can be produced. A logbook may record important incidents or comments from an event so that they may be accurately recalled. These entries should be made in real time with a clear indication of students name and the date of event. A sequence of this type of records may be called as a running record.

6.1.3 Typical Observation Formats

Narrative recording
This format involves simply noting the steps of the event’s occurrence. This is a qualitative method which is not countable. It provides general details that what happened.

A-B-C
This format requires dividing paper into three columns. In first column “A” is for antecedent. This column consists of what happened before the event. The column “B” is for behavior. It provides information about the students’ action. The column “C” stands for consequence that means that what is the outcome of the behavior. This consequence may become antecedent for the next behavior and so on a sequential record can be maintained

<table>
<thead>
<tr>
<th>A-B-C Data Sample (Source: R. Van Acker)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student:</strong></td>
</tr>
<tr>
<td><strong>Activity:</strong></td>
</tr>
<tr>
<td><strong>Time of day:</strong></td>
</tr>
<tr>
<td><strong>Antecedent</strong></td>
</tr>
</tbody>
</table>

117
Scatter Plot
Scatter Plot is a tool that determines the time and space for the occurrence of the behaviour. The sample format is described below.

Functional Assessment Scatter Plot
Student: __________________________ Grade: ______
School: __________________________
Dates of observation: _______ to _______ Observer(s): ______
Target behavior(s):

<table>
<thead>
<tr>
<th>Activity</th>
<th>Time</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: Touchette, Macdonald & Luce

Task Engagement and academic participation
This format helps to collect the data of the students while they are in the mid of any task so that their behaviour can be recorded at that time and their willingness to indulge in the required activities during the session. This format is comparative in its nature. It provides the data of other students with the subject students to compare with.

Task Engagement – the observer will need to develop an operational definition for what they deem to be task engagement or on-task behavior.

Task Engagement:

...........................................................................................................................................................................
...........................................................................................................................................................................
...........................................................................................................................................................................

The observer should employ momentary time sampling – observe the student at the last second of a pre-selected time interval and record whether the student is engaged in the task/activity (+) or not engaged (-). The observer must specify the time interval employed (e.g., 15 sec., 1 min., 3 min., etc)
Task Engagement (Employ Momentary Time Sampling - Interval Length = ____ seconds)
Nature of the task:

(Target Student)

Nature of the task:

(Target Student)

Nature of the task:

(Other Student)

Nature of the task:

(Other Student)

Academic Participation

It examines the extent and the manner in which the student displays willingness and/or the teacher activities. The observer simply places an (x) in the appropriate involves the student in oral participation during classroom box for each opportunity provided and response given. For example, if the student volunteers to answer a question, is called upon by the teacher, responds correctly, and is praised – the box would be marked as shown.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Volunteers</td>
<td>X</td>
</tr>
<tr>
<td>Called on</td>
<td>X</td>
</tr>
<tr>
<td>Called out</td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td>X</td>
</tr>
<tr>
<td>Consequence (+/-/0)</td>
<td>+</td>
</tr>
</tbody>
</table>
**Academic Participation:** (Employ Event Recording - Tally marks)

(Target Student)

<table>
<thead>
<tr>
<th>Volunteers</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Called on</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Called out</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consequence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+/-/0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Other Student)

<table>
<thead>
<tr>
<th>Volunteers</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Called on</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Called out</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consequence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(+/-/0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: This form may be photocopied for noncommercial use only. Copyright © Van Acker & Associates 1998

**Event Recording**

It is the count of the occurrence of a specific behaviour in a specified time period. It is the most easy and accurate way of data collection. The under consideration behaviour must have a well defined start and end and time intervals should be same. Each event could easily be distinguished from the other.
<table>
<thead>
<tr>
<th>Intervals</th>
<th>Behavior</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attentive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooperating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demanding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dependent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disruptive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exploring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Helping Others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imitating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In Transition/Changing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inappropriate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recalls Routine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initiates Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leading/Demonstrating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Making Choices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On-looker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requesting Help</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncooperative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uninvolved/Wandering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unsafe Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using Materials Appropriately</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using Appropriate Manners</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waiting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Deriving Results from the Observation by Developing Rubrics**

A rubric can be termed as a scoring tool that exactly measures the performance expected during a task. Rubric subdivides a task into a number of components and clearly describes the characteristics of that specific component with variable levels of mastery. Rubrics are proved to be useful in the cases of individual performance, essays, exam questions, story writing, projects, worksheets etc.
Rubrics are very useful for the assessment of students’ performance. They provide a scale to measure the students’ ability to solve some task. As discussed before that teachers record data during the observation session and then they develop a rubric for this purpose. Rubrics subdivide the task into many components and assign level of score for each component. A criterion is then developed to assign scores for each component. On the basis of the observation sheet teacher starts assigning scores to each component according to the set criteria, Hence a full scorecard develops which provides true picture of student performance in a specific task. Let us elaborate this phenomenon with the help of an example.

A teacher assigned her students to write a story according to the environment. Instructed them to observe their surroundings and develop a story. While development of story she started to take the observation notes and recorded them. These recordings included the students’ interest, observation, the structure of story, the use of language etc.

After recording this data the teacher will develop a rubric that will subdivide the task into many components like interest, observation, structure, plot and language. Then she will supposedly assign marks to each section and then she will on the basis of her observation sheet assign the obtained marks for each student for each component and at the end by calculating the total number of marks. We will be able to make decisions about the performance of the specific student. This is how rubrics work and prove themselves helpful in deriving the results from the performance of students.

Sample Rubric

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Portfolio Requirement</strong></td>
<td>Poor Portfolio is missing more than 4 minimum requirements as stated in the syllabus.</td>
<td>Fair Portfolio is missing 3 minimum requirements as stated in the syllabus.</td>
<td>Good Portfolio meets all minimum requirements as stated in the syllabus.</td>
<td>Exceptional Portfolio meets all minimum and above requirements as stated in the syllabus.</td>
<td></td>
</tr>
<tr>
<td><strong>Creative use of Technology</strong></td>
<td>Poor No use of graphics, Internet resources, photographs, sound and/or video to enhance Portfolio and reflective statements.</td>
<td>Fair Little use of graphics, Internet resources, photographs, sound and/or video to enhance Portfolio and reflective statements.</td>
<td>Good Some use of graphics, Internet resources, photographs, sound and/or video to enhance Portfolio and reflective statements.</td>
<td>Exception Good use of graphics, Internet resources, photographs, sound and/or video to enhance Portfolio and reflective statements.</td>
<td></td>
</tr>
</tbody>
</table>
### Artifacts

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>Most artifacts and work samples are unrelated to the purpose of the course and portfolio.</td>
</tr>
<tr>
<td>Fair</td>
<td>Few artifacts and work samples are related to the purpose of the course.</td>
</tr>
<tr>
<td>Good</td>
<td>Most artifacts and work samples are related to the purpose of the course and portfolio.</td>
</tr>
<tr>
<td>Exceptional</td>
<td>All artifacts and work samples are clearly and directly related to the purpose of the course and portfolio.</td>
</tr>
</tbody>
</table>

### Organization & Writing

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>The text has many errors in grammar, capitalization, punctuation and spelling requiring major editing and revision.</td>
</tr>
<tr>
<td>Fair</td>
<td>The text has errors in grammar, capitalization, punctuation, and spelling requiring editing and revision. Easy to read and navigate.</td>
</tr>
<tr>
<td>Good</td>
<td>The text has a few errors in grammar, capitalization, punctuation, and spelling requiring editing and revision. Easy to read and navigate.</td>
</tr>
<tr>
<td>Exceptional</td>
<td>The text has no errors in grammar, capitalization, punctuation and spelling. Easy to read and navigate.</td>
</tr>
</tbody>
</table>

### Reflections

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor</td>
<td>Considerable difficulty in expressing reflections demonstrates level of students' progress and knowledge development. Reflections incorporate the what, so what and now what in artifacts.</td>
</tr>
<tr>
<td>Fair</td>
<td>Difficulty expressing and reflecting demonstrates level students' progress and knowledge development. Reflections incorporate the what, so what and now what in artifacts.</td>
</tr>
<tr>
<td>Good</td>
<td>Most reflections demonstrate students' progress and knowledge development. Reflections incorporate the what, so what and now what in artifacts.</td>
</tr>
<tr>
<td>Exceptional</td>
<td>All reflections demonstrate students' progress and knowledge development. Reflections incorporate the what, so what and now what in artifacts</td>
</tr>
</tbody>
</table>

### 6.1.4 Making Decision on the Basis of Observation Results

After going through the complete process of observation and recording data we apply rubrics to know about the measureable scale of students’ performance. Rubrics clearly define students’ performance in each component of the task as well as for the whole task. As we know that these derived results are the ultimate consequence of the process of keen observation, so we are now ready to make decision on the basis of these observatory results. By taking the previous example
in the view we can assess the level of mastery of the student in each component and make decisions as if an individual students’ needs improvements in observations or creating the plot of story. In this way we can guide every student according to his/her needs. And this process helps in making micro level and macro level decision for a class, a school or any other educational facility. We can say that process of observation is the base for making decision that helps in improving the students’ performance in a specific level of education.

6.2 CLASSROOM PROJECTS

From a couple of decades, several efforts have been made to improve teaching and learning processes in classrooms. The efforts include implementation of child-centered curricula, employing professional standards for teachers and the use of technology etc. Colley (2005) stated that despite all these reforms, the teaching and learning process in classrooms is practiced still through chalk-talk method. The most common instructional activities in classrooms are lecture and discussion. Learning to explain ideas and analysis/synthesis of arguments based on concrete evidences have been given very less emphasis at all levels resulting students leading towards rote memorization. In order to address this issue, classroom projects have been used as alternative techniques. Classroom projects are the experience centered teaching-learning activity. The main focus of classroom projects is socializing the child and developing the problem solving ability.

The concept of project was introduced by W. H. Kilpatrick in 1918. His emphasis was on the purposeful learning activity and problem solving skills among the students on the basis of their needs, interest, attitudes and abilities. He was influenced by the John Dewey’s learning by doing principle. It was because of the idea that projects support learning to be related to the life situation of the pupils.

6.2.1 Definition of Project

Project is defined as a whole-hearted purposeful activity proceeding in a social environment. It is a willingly taken responsibility which involves constructive effort to achieve the desired objective. A project is an in-depth study of a particular topic that one or more children can undertake for a certain period of time. Projects are multifaceted tasks involving exigent questions to be answered, demanding from learners to design, solve problems, make appropriate decisions and investigate to explore.

Classroom projects provide opportunities to students to work autonomously which is relative over a certain periods of time and the end result of projects is
culmination of assigned or undertaken task in the form of realistic products. Classroom projects make classrooms function like mini experimental stations, research laboratories and inquiry agencies. During classroom projects, students pose questions and conduct extended studies to find answers to their questions within the context of a unit, curriculum or program of study. Teachers act as facilitators, mentors, resource persons, advisers, listeners, learners and leaders in classroom. They work with students to identify projects that they are interested in and create learning environments that allow them to collect resources, plan, implement, evaluate and report on their projects. The students decide what to learn, how to learn, the time required to learn and how to document and report their own learning. When students are given responsibility for their learning and hold them accountable, they are more likely to take it seriously and rise up to the challenge than if they are spoon-fed.

**Types of projects**
According to Kilpatrick there are four types of projects.

**Constructive project**
Practical or physical tasks such as construction of article, making a model, digging the well and playing drama are done in this type of projects.

**Aesthetic project**
Appreciation powers of the students are developed in this type of project through the musical programmes, beautification of something, appreciation of poems and so on.

**Problematic project**
This type of project develops the problem solving capacity of the students through their experiences. It is based on the cognitive domain.

**Drill project**
It is for the mastery of the skill and knowledge of the students. It increases the work efficacy and capacity of the students.

**Other types**

**Individual and social (Group) projects**
In individual project students solve the problems by their own according to their interest, capacity, attitude and needs. It develops the problem solving qualities individually and not the social qualities. On the other hand in group projects the
problem is solved by the group of pupils in the class. Here the social, citizenship qualities and synergism are developed.

**Simple and complex projects**

In the simple projects the students are completing only one work at a time. They are also focus the work in one subject or one area only. It gives the deep information about the project in one angle. The students get knowledge about the problem deeper and broader. In the complex project the students carry out more than one work at a time. They focus on the work in various subject and angles. Here the students get the knowledge about the work in various activities and dimensions.

### 6.2.2 Tasks Versus Tests

It is common observation that teachers, educators, educational managers, administrators and students use educational terms without differentiating their meaning and applicability. Tasks and tests are also being used in educational dialogues, discussions and in written materials. Tasks are such activities where learners have multiple opportunities to practice the required knowledge and skills over a longer period of time. Working on tasks, students are allowed to learn and revise content multiple times. For example, homework is an example of task to be completed at home as an extension of learning activities of classroom. As end result of the tasks, teachers can have an in-depth overview of students’ progress on the tasks performed.

Conversely, tests are designed to be completed in a given period of time under specific conditions. In tests, specific skills and knowledge about a certain topic is required in the given time. The purpose of the tests is to find out the extent to which students have got mastery in a particular content area. Tests also serve as means to know how well students are doing in topics under study. Through tests, teachers can have a picture of the strengths and weaknesses among the students. In tests, once attempted there is no chance for revise and correct the mistakes until marking.

**An example of tests**

With the help of given verbs, fill in the blanks by writing correct form of verb to complete the sentences in simple past tense. (read, send, complete, go, arrange) (5marks)

1. Yesterday, Salma ________ to academy and studied science.
2. Last evening, Iram ________ an interesting book on History of United Kingdom.
3. They________ a wonderful programme this morning.
4. Shahina ____________ an email to me last week.
5. The students ____________ their homework.

**An example of a task**

Planning for a party, the learners are expected to carry out the following.

1. Discuss and decide what they need for preparation.
2. Select a place or venue where the party will be held
3. Prepare for the party
4. Write invitations

Activity 1: Develop a test for your interesting subject area.
Activity 2: Design a task to be completed by the students related to your subject area.

**6.2.3 Characteristics of a Project**

In the following, characteristics of a project are given.

**Purposefulness**
The projects are purposeful and have some objectives which energize students to work diligently. Without purposefulness any learning activity will be wastage of time and energy.

**Utility**
Projects are useful for the students and the society. There are opportunities for students to develop some values. From projects, the students as well as the society get the benefits a lot.

**Freedom**
The students are free to select the topic and execute the work according to their own pace, interest, attitude and capacity. The teacher just acts as a guide and gives guidelines to execute the project.

**Activity**
Project means the purposeful activity, at the end of the project the students gain knowledge, skills and values through their activity. It is based on the principle of learning by doing.

**Reality**
Projects are realistic and related to the life situation of the students and the society. The students develop ability to complete the projects which are realistic and time bounded. Imaginary problems are not taken up in the project.
6.2.4 Indicators of a Successful Project

Classroom projects are appraised in terms of context, input, process and output. The following indicators show the success of any classroom project.

**Contextual Indicators**
In terms of context, the indicators of successful projects are;
1. A successful project meets the needs of the students such as project is according to age, mental capacity, interest, paying capacity of students if applicable etc
2. Culturally not sensitive
3. Helpful in resolving local community issues
4. Challenging that means no so easy and not so difficult
5. Supportive in teaching and learning

**Input Indicators**
Regarding input during a classroom project, the indicators of a successful project are given below.
1. Use of resources that facilitate learning
2. Financial, material and human resources are utilized according to the plan
3. Appropriate instructions/guidelines are followed and records are maintained
4. All inputs are aligned with the objectives of the project
5. Ensure the inputs to be available by the end of the project

**Process Indicators**
A project will be considered successful in terms of process when the following indicators are evident.
1. All the activities are carried out according to the plan
2. Followed standards as designed
3. Quality of the activities are above benchmarking
4. Pace of the working is according to the timeline
5. All the stakeholders are satisfied with the processes of the project

**Output Indicators**
In terms of outputs, the indicators of a successful project include;
1. The outputs in the form of students’ learning, reports, presentation show the attainment of the objectives.
2. Project reveals increasing knowledge, skills and experience
3. Performance of the students after project has enhanced which were measured through tests and examinations
4. There is a change recorded in the learning place/classroom situation
5. Satisfaction of all stakeholders on the project after completion
6.3 PHASES/STEPS OF A PROJECT

Following are the phases/steps followed while carrying out a project.

Creating a Situation
In the first step teacher creates the proper situation to the students in the class. He/She puts up the knowledge about the project, procedure, steps and uses to the students. After that he/she should motivate the students through conversation about the day to day life problems to the students.

Selection of the Problem
Then the teacher helps the students to select the problem and guide them. Here the students are having freedom to choose the topic or problem based on their interest and ability. Before choosing the topic the principles should be taken in to account.

Planning the Project
The teacher discusses with the students about the problem in various angles and points. He/she should create the situation to discuss with the students and they are allowed to talk freely and openly. After the free expression of the students’ opinion about the problem, the teacher writes down the whole programme of action stepwise on the blackboard. The grouping is made by the teacher based on the interest and ability of the students.

Executing the Project
The students start working in this stage. They collect the relevant information/data and materials. The teacher gives the time and right to the students according to their own speed, interest and ability. If need arises the teacher will provide the necessary help and guidelines to the students. He/she demands the groups to complete the project in the given time.

Evaluating the Project
Here the students evaluate their project. They determine whether the objectives were achieved or not. After that they criticize and express their feeling about the project freely. They discuss the planning, selecting the task, execution and the entire process in the class.

Reporting and Recording
It is the last step of the project method in which each and every step of the work is reported. The reported stuff is recorded in a certain order in a book form. The record is useful for the further use and future reference about the project. It
reveals many ideas about the concern project. The book formatted report is submitted to the teacher at the end.

6.3.1 How Can Projects be implemented in a Classroom?

There are different strategies teachers can use to implement projects in their classrooms. One of effective way to implement project is to tie it to a standard, unit, curriculum, or a program of study. Colley (2005) gives some steps which could be followed while implementing classroom projects:

1. Conduct a pre-assessment of the students regarding the prior knowledge, skills and attitudes about the topic, theme or unit of the project. Discuss the advantages, procedure, steps and principles of project work with the students. Give responsibility and accountability to the students to generate questions to be worked on.

2. Prompt the students about the importance of collaborative work in carrying out a project based on the principle that more heads are better than one head on a single question to be addressed. Discuss the advantages and disadvantages of working in groups.

3. Suggest the students while working in groups raise questions, argue and justify each other’s arguments, show the behavior of agreeing and disagreeing to prove the passed statements. Bring in to account the strengths and weaknesses of project work during execution.

4. Form heterogeneous balanced groups of students in terms of gender, academic performance, socio-cultural backgrounds and so on.

5. Make it sure that each group will formulate a question to be investigated within their own context. Each group will develop a plan regarding project work and each group will present their work after completion of the execution and separately submit the report.

6. Develop a timeline for each group and the teacher needs to take that in his/her own record for monitoring.

7. Conduct a post assessment at the end of the process. The process of project is complex but the key is to have students generate purposeful, doable, relevant and interesting questions, implement them within the available timeline and reflect on their own learning.

6.3.2 Assessment of a Project and Use for Evaluation

While assessing a project and the information collected through assessment is used for evaluating the project, it is important to raise the following questions by the assessor: What problems are being tried to be solved? Prior to assess the project, the assessor needs to be clear about what are expected to be achieved that
means what objectives have been designed. While reviewing the situation in which projects have been carried out, the assessor needs to find out what are the problems being solved? What is making a change? What needs to be changed and why? What factors have been identified which caused the problem? What plans have been designed to bring about changes? What and how actions have been taken? What resources have been used? What indicators have been listed down that will show the achievements during the project? How has been progress measured? The answers to these questions will provide a framework for evaluation. Gathering data or evidences is the key step of the assessment procedure. The evidences which show the success of the projects allow the assessor to find out the extent to which objectives have been achieved. The evidences in different forms need to be collected such as numbers, samples of work done during the project work, reflective notes by the group members, success stories, opinions, views and experiences, pictures, difference between the scores of pre-assessment and post assessment etc. the assessor needs to establish a baseline that will lead to indicate the starting point and small milestones achieved during the projects. It should be ensured that the evidence or data gathered in meaningful and is in concrete form. Once the evidences are collected, then they need to be analyzed qualitatively and quantitatively to report the findings. These reports then are used in evaluation of the project.

6.4 METHODS TO CALCULATE RELIABILITY

Double Marking/Inter-rater Reliability
In this method to calculate reliability, two or more independent raters rate the test. The scores are then compared to determine the consistency of the raters’ estimates. One way to test inter-rater reliability is to assign each rater score each test. For example, each rater might score items on a scale from 1 to 10. Then the correlation between the two ratings is found to determine the level of inter-rater reliability. Another means of testing inter-rater reliability is to have raters determine which category each observation falls into and then calculate the percentage of agreement between the raters. So, if the raters agree 8 out of 10 times, the test has an 80% inter-rater reliability rate (Swain et al, 2000).

Split-half Method
Linn and Gronlund (2000) shared that the split-half method of determining internal consistency employs single administration of a test on a sample of pupils. The test scores are then divided into two equivalent halves and correlation for the test scores of both halves is calculated. The test can be divided into even numbered items such as 2,4,6…, in one half and odd numbers such as 1,3,5,…., in
another half. Then the scores of both the halves are correlated by using spearman brown formula. The formula is given below.

6.5 SELF ASSESSMENT QUESTIONS

1. Enlist the different types of techniques and their role in education system.
2. Define the term reliability and elaborate the importance and scope of reliability of a test.
3. What factors should be considered in deciding whether subjective or objective type questions should be included in classroom tests?
4. Write down your preferences of selecting Multiple Choice Questions rather than True-False items.
5. Describe the various types of reporting test score by giving examples from our country context.
BIBLIOGRAPHY


133
DEVELOPING ALTERNATE TECHNIQUES-II

Written By: Hassan Arslan Khalid
Reviewed By: Muhammad Zafar Iqbal
CONTENTS

Introduction .................................................................................................................. 137
Objectives .................................................................................................................... 137
7.1 Oral Questioning .................................................................................................. 138
   7.1.1 Purpose of Questioning (e.g. feedback for improving teaching and
         learning) .............................................................................................................. 139
   7.1.2 Guidelines for Questioning ........................................................................... 142
7.2 Peer Appraisal ...................................................................................................... 144
   7.2.1 Guess Who’ Techniques .............................................................................. 144
   7.2.2 Socio-metric Techniques .............................................................................. 145
7.3 Interview ............................................................................................................... 145
   7.3.1 Types of Interviews ...................................................................................... 145
   7.3.2 Interview Strengths and Weaknesses .......................................................... 146
   7.3.3 Use of Interview for Students’ Evaluation ................................................... 148
7.4 Portfolio Assessment ............................................................................................ 148
   7.4.1 Types of Portfolios (increasing breadth, increasing depth) ......................... 149
   7.4.2 Steps in the Portfolio Assessment Process .................................................. 150
7.5 Computer-Assisted Testing and Future Trends in Assessment .......................... 153
7.6 Self Assessment Questions .................................................................................... 156
Internet Sources ......................................................................................................... 158
INTRODUCTION

Assessment is at the heart of learning and decisions about the modes and methods of assessment can exercise a profound influence on students’ experience, how they perceive their courses, how they relate to each other and to their teachers, and how they prepare for the world of practice beyond the university. Teaching, learning and assessment processes that strengthen students’ engagement with ideas, that develop identity, and that build relationships within communities of learning depend on multiple modes of communication. In an age of mass higher education, the continuing and expanding use of oral assessment as part of a carefully designed mix of assessment types provides rich opportunities for student engagement as well as opportunities for enhanced judgments about student achievement. Oral assessment raises many challenges, it also offers considerable rewards for teachers and students alike.

OBJECTIVES

After studying this unit, the students will be able to:
1. describe purpose and guidelines for oral questioning as an assessment technique.
2. elaborate different techniques used in peer appraisal.
3. justify the interview as an assessment tool by discussing strengths and weaknesses of interviews
4. discuss process of portfolio assessment technique.
5. explain future trends in assessment of students.
7.1 ORAL QUESTIONING

Oral questioning is a valuable form of assessment that is most commonly used in the classroom. In fact, it is without a doubt an important feature of practically all teaching to say the least. Questioning is used to elicit information or concepts learned from the students in order to assess if they have mastered the concepts. Oral questioning is a type of formative assessment, that is, it can be used throughout a lesson. The teacher elicits information from the student and evaluates the information. Oral questioning can be done to assess and activate prior knowledge, and during and after a lesson to assess comprehension and mastery of a concept.

Questioning should only be done when used to assess what is directly relevant to the task. Questions can be about what did or didn’t happen, seek an explanation for a particular practice, check for understanding of underlying principles and challenge a practice. There are many more purposes for the types of questions that teachers ask. The main goal of questioning for assessment is for the teacher to be able to understand and evaluate what the students know and don’t know. For example, in a Spanish 1 classroom the teacher may ask a student why verbs are modified. If the student answers the question correctly then we know he/she has at least recalled the information and somewhat mastered the concept. If the student needs to be scaffolded then the teacher may wish to reteach why verbs are modified to the class and continue with questioning. In a more advanced Spanish class students can answer questions that require higher thinking skills.

The importance in questioning lies in that teachers need to use various questioning strategies to set up students for success. For example, teachers should plan questions as they prepare their lessons, considering the instructional goals and emphasizing questions that reinforce them. Teachers should use vocabulary that students can understand, ask clear and specific questions, and be sure to incorporate questions from all levels of Blooms Taxonomy. Questions should also be constructed to connect important concepts and give the teacher feedback above all else. It’s also important to understand that if students aren’t answering the question as the teacher expected then the question needs to be rephrased. The teacher also needs to understand that students need time to think about the answer, so wait time should be allowed when eliciting a response from a student.

Oral questioning in the classroom is the most common form of assessment. Without a doubt questions should play an important role in every classroom, language arts, math, social studies, foreign language, etc. It requires
lots of mental preparation on the teachers’ behalf, and when the questions are set up properly can be effective in assessing the students.

Oral questioning as well as oral review techniques can be used for a variety of purposes, which include, but are not limited to:

- Introducing, summarizing, or reviewing a lesson
- Clarifying points previously made
- Bring up points omitted
- Bringing reading assignments into focus
- Developing in students new insights
- Promoting students’ understanding
- Developing students’ attitudes and values
- Teaching students to use ideal rather than to simply memorize them

Oral questioning can provide important evaluation information. Students’ preparation for lessons can be tested. Preliminary questions during the lesson introduction can serve as a pretest of students’ knowledge level. Also, using review questions during the lesson can provide immediate feedback on how well students are progressing.

Why assess orally? There are many reasons for choosing to assess our students orally. Here are seven commonly used reasons for opting for oral assessment, either on its own or in combination with other modes of assessment.

7.1.1 Purpose of Questioning (E.G. Feedback for Improving Teaching and Learning)

**Purposes of Oral Questioning**

The primary purpose of oral questioning is to stimulate the students to think. Oral questioning also provides you with a practical means for establishing the level of instruction. Students may vary greatly in the quantity and quality of background knowledge they have acquired through previous training and experience. You must determine the level of achievement of the students before proceeding with the presentation of new subject matter. Although you may use a pretest or a questionnaire for this purpose, the quickest and simplest means is a series of oral questions. Oral questioning has three other important purposes:

First, it arouses interest in the subject matter.
Second, it focuses attention upon a particular area of the subject matter.
Third, it drills students on subject matter they must recall precisely, such as correct terminology, functions of parts and safety precautions.
Use questions to achieve the following benefits:

- Discover each student’s interests, abilities and depth of knowledge.
- Arouse student interest in the subject matter of the lesson.
- Stimulate discussion and keep it closely tied to the subject matter.
- Review and summarize important points.
- Test students’ knowledge of what the lesson has covered and check the effectiveness of the instruction.

**Characteristics of A Good Oral Question**

Questions that are poorly worded, vague in meaning, or ambiguous will frustrate both you and the students. Students who do not comprehend the true meaning of poorly phrased questions will hesitate longer than usual and then give uncertain answers. You may feel dissatisfied with the answers and reprimand the students for their lack of attention and understanding. The students, knowing that they have answered unsatisfactorily through no fault of their own, may lose enthusiasm and withdraw from active participation. You can avoid frustrations of this kind by planning your questions well in advance as well as carefully choosing and arranging words and phrases. The construction of good oral questions requires three considerations; level of instruction, use of interrogative and clarity of meaning. A good review question should have the following characteristics:

- Be concise, including only one idea
- Be short enough for students to remember
- Be timely, interesting, thought-provoking and relevant to the lesson being taught
- Be stated in language familiar to students
- Be stated to stress the key points of a major lesson topic
- Be stated to require more than a guessing response
- Be stated to require more than a simple yes or no answer
- Be stated in such a way that it does not suggest the answer

Your method of reviewing can be classified according to the level of knowledge required for the correct response. Bloom’s taxonomy of educational objectives is built on a progression of complex levels. At the lowest level, students are asked to only recall or recognize the correct responses from memory. Then the levels increase in complexity to comprehension, application, analysis, synthesis and evaluation.

**Seven reasons for using oral assessment**

1. The learning outcomes demand it
2. It allows probing of the students’ knowledge
3. It reflects the world of practice
4. It improves learning
5. It suits some students
6. The meaning of questions can be clarified
7. It helps to ensure academic integrity

**Level of Instruction**

In asking questions, use simple words, correct grammar and complete sentences. Use words the students know and understand. As the course progresses, introduce new terms and more technical phraseology. Ask questions at times that suit your presentation of course material. Plan questions that require students to think before answering. Don’t use questions that give away the answer or that students can answer with a simple yes or no.

**Use of Interrogative**

Use the interrogatory word or phrase at the beginning of your question so that students know immediately when you are asking a question. Let’s consider two examples where this is not done: (1) The two sizes of fire hose most frequently used in the Navy are what? and (2) You can determine whether or not explosive vapors are in a compartment by what means? Questions worded in this way handicaps the students in at least two ways. First, the students are expecting a declarative statement, not a question. Second, they cannot identify the meaning of the question until the final words are spoken. Note the improvement in these same questions when the interrogatory word or phrase is placed at the beginning: (1) What are the two sizes of fire hose used most frequently in the Navy? and (2) By what means can you determine whether or not explosive or toxic vapours are in a compartment?

**Clarity of Meaning**

Avoid the use of catch or trick questions as a teaching device, especially for beginners. Make sure the wording of the question conveys to the students the true or intended meaning. The students must understand what you want, regardless of whether they know the correct answer. “Where are storm warnings flown aboard ship?” is a good question; but “Where are storm warnings flown?” fails to indicate what point is being tested. Make your questions brief and limit them to one thought. To include too many factors in a single question confuses the students. Ask well-stated, clearly understood questions in a normal conversational
tone as part of the lesson. After each lesson, reevaluate your questions in light of how the student responses contributed to better learning.

**Types of Oral Questions**

Learn to use oral questions throughout the lesson. Use them in the introduction to create interest and focus attention on the subject matter and during the lesson presentation to ensure student understanding. Then use them at the end of the presentation for review and drill purposes. Feel free to ask factual, thought-provoking and interest-arousing questions as often as you choose. Other types of questions may serve one or more useful purposes if used sparingly but may prove ineffective if you use them too often.

**Factual Question**

The factual question asks for specific information; for example, “When was the first U.S. nuclear powered submarine built?” Although the primary purpose of the factual question is to help students memorize facts, it may, under certain conditions, have important secondary.

**7.1.2 Guidelines for Questioning**

There are three sub-levels of knowledge involved in comprehension. The levels are translation, interpretation and extrapolation. In each of these sub-levels, the student is expected to expand his/her thinking past the level of simply recalling information.

The application level requires students to solve practical problems through the selection and use of ideas, principles and theories. After application, the analysis level will ask students to break a whole down into its component parts and to determine the relationship between the parts. The synthesis level requires students to put together parts and elements to form a new whole or pattern. Finally, the evaluation level requires students to make judgments based on specific criteria rather than opinions.

When employing oral review questioning techniques, there are certain guidelines that instructors need to consider. This is called the “six shoulds”.

1. Review questions should be distributed among class members so that each student has the opportunity to participate.
2. Review questions should be asked in a normal conversational tone, loud enough for all class members to hear.
3. Review questions should be presented in a logical sequence.
4. Students’ responses should be repeated for special emphasis or clarity.
5. Students should be encouraged to go beyond the first answer, expanding and corroborating on what others have said.
6. A review question should be directed to a particular student after the question has been asked to encourage other students to formulate answers.

Some disadvantages of oral assessment

Undue anxiety
Some anxiety can be beneficial in oral assessment, but anxiety that interferes with a student’s performance will not give a true indication of his or her ability. Anxiety may be a special impediment for students with particular mental health problems. Practising presentations in class and providing rehearsals for vivas may help. Sometimes a student who experiences undue anxiety may need to be accommodated through alternative arrangements for their assessment.

Hearing or speech difficulties
Students with hearing or speech impairments may also require some adjustment to the assessment process.

Time
Oral assessment can be time-consuming, which becomes particularly problematic with larger classes. On the other hand, many forms of oral assessment can be quite short, and marking can occur very quickly at the end of the assessment. For example, Roberts describes a Geography viva which takes 10–15 minutes per candidate, including paperwork (Roberts, n.d.).

Lack of anonymity
Examiners inevitably know whom they are examining.

Bias
Concerns are sometimes expressed that examiners may be influenced by students’ dress, gender, ethnicity or educational background.

Novelty
The form of oral assessment being used may be unfamiliar to the student.

Recording
Many universities, and good practice, require us to keep a record of the assessment for future reference in case of appeal. Making and storing audio or video recordings can be difficult to arrange.
Articulateness vs knowledge
Examiners can mistake a student’s articulateness for knowledge.

7.2 PEER APPRAISAL

Peer appraisal is a core part of our quality control it provides a mechanism for:
- monitoring the quality of our educational provision and performance of the instructor(s),
- professional recognition of the variation of modules in innate difficulty and presentation,
- identifying alternative approaches to delivery of material,
- the identification of good practice.

In peer assessment, a collaborative learning technique, students evaluate their peers’ work and have their work evaluated by peers. Often used as a learning tool, peer assessment gives students feedback on the quality of their work, often with ideas and strategies for improvement. At the same time, evaluating peers’ work can enhance the evaluators’ own learning and self-confidence. Peer involvement personalizes the learning experience, potentially motivating continued learning.

When used in grading, peer assessment can give the instructor needed information on student performance. Especially for large online classes, it may allow inclusion of assignments where students’ creative work could not be graded reliably through automation or efficiently by teaching staff.

Peer assessment techniques vary considerably and are often best understood through example. To give effective, valid and reliable feedback to fellow learners, students need clear guidelines, training on assessment criteria and scoring rules, and practice with examples. Before students are ready to give feedback to others, their assessments should be compared to staff-grading of the same examples for quality assurance.

7.2.1 Guess Who’ Techniques

A type of personality rating device used chiefly in schools. Students are given short word pictures depicting a variety of personality types and are directed to identify the classmates whose personalities seem to correspond most closely to those descriptions.

The guess-who technique was developed by Hartshorne and May (1929). In this technique a student is asked to read each descriptive statement and then write
down the name of the student who best fits that description. The student may write more than one against each statement and allowed to write his own name.

7.2.2 Socio-Metric Techniques

The term sociometry is defined as the measurement of the social relationships that exists among the members of a group. Sociometric technique attempt to describe attractions or repulsions between group members by asking them to indicate group members and by asking them to indicate whom they would select or reject in various situations (Koul, 1998).

Uses of Sociometric Techniques: It is used to study Social adjustment, Group dynamics, Learning, Motivation, and Discipline.

7.3 INTERVIEW

Interviewing involves the interaction in which an interviewer collects information from students with a sequence of questions and listens for answers. This kind of interaction can be a rich source of information to inform the teacher about how the student understands concepts and use procedures they learned from the course, and provides valuable information and directions for the teacher in modifying the course for improvements.

7.3.1 Types of Interviews

Structure of an Interview
Although interviews can be conducted over telephone or other forms of media, it is usually done face-to-face.

There are two main types: Structured and Unstructured interviews
Structured interviews are composed of a series of well-chosen questions which are designed to elicit a portrait of a student's understanding about a concept or set of related concepts. To explore the topic more deeply, probe questions are commonly used to follow up those pre-planned 'main' questions. These probe questions are usually not formally designed ahead of the interview. Probe questions are usually formed according to the responses and answers given by the interviewee. When the interviewer finds the responses/answers are worthy to be explored more deeply, they would ask their interviewees to elaborate the content further. This approach ensures that interviewer and interviewee have thoroughly finished exploring one topic before moving onto another. Variations of structured interviews include: Instances Interviews, Prediction Interviews, Sorting
Interviews, Problem Solving Interviews etc. Instances Interviews: a student is presented with a specific set of examples and counterexamples of the concept of interest and is asked to identify which cases are examples of the concept, and then to explain that decision.

Prediction Interviews: students are required to anticipate an outcome of a situation and explain or justify that prediction.

Sorting Interviews: the student is presented with a group of objects and asked to sort them according to specific instructions.

Problem Solving Interviews: a student is asked to attempt to solve a problem while thinking aloud, explaining as much as possible about what she is doing, why she is doing it, and what her symbols and actions mean?

Unstructured interviews are used when the interviewer wants to let the interviewee have complete control over the content of the interview. The interviewer usually prepares one or two questions to start off the interview. Only probe questions would then be used for the rest of the interview for further elaboration on a topic.

7.3.2 Interview Strengths and Weaknesses

Advantages of an Interview

- **In-depth information:** In contrast to the set of fixed questions used in surveys, interviews allow the probing of deeper information from the interviewee, asking for elaboration and examples, discovering new relationships and the interviewer can modify questions as the interview proceeds.

- **Rapport and trust:** Good interviewers can establish rapport and trust with the interviewee, which can also elicit more honest and more sensitive information than surveys.

- **Level of understanding by learners:** Structured interviews enable instructors to examine the degree of understanding a learner has for a concept or closely related group of concepts, and to focus on how their instruction is interpreted and internalized by the learner.

- **Guides improvement:** When well-administered, interviews can be a powerful type of formative assessment to guide improvement in courses and teaching methods, as well as enabling teachers to understand the typical difficulties faced by students in the course.
Disadvantages of an Interview

- **Time consuming:** Every interview would need to take approximately 30 to 90 minutes to finish. Also, as the nature and quality of probe questions and follow-up questions will determine the usefulness of the interview, interviewers have to take a certain amount of time planning and designing an informative interview. In addition, as it takes some time to finish the interviewing process and data analysis, it might be quite a while before interviewees can receive their feedback.

- **Bias from interviewers:** It is possible that sometimes the interviewers might somewhat bias the nature of the interview data and thus the results, through their verbal and nonverbal reactions, and their choice of probe questions when interacting with students during the interviews.

- **Bias of interpretation:** If interviewers are inexperienced, improperly trained or careless, even though the interview results have been "coded" (content analyzed), there are still possibilities for them to be biased when interpreting and summarizing the results.

- **Subjectivity of interviewees:** Information obtained by interviewers is based on the perceptions, knowledge, and words of interviewees, rather than objective and behavioral data. Interviewers have to rely on their interviewees words regarding the accuracy of the information.

How to design a good Interview Assessment?

- Try to make the student feel relaxed and comfortable during the interview, because interviews can generate the most fruitful sharing when a trustful rapport is established.

- Practicing can help to ensure that the interview can be finished in a reasonable amount of time (normally less than an hour).

- Carefully select the sample of students for interview so that the group can represent all students who may have different levels of interest and ability.

- Ensure all the necessary equipments for the interview are well-prepared, such as interview protocol, audio and video recorders etc.

- Try to make the interview group as small as possible, or conduct it individually; this can best facilitate in-depth sharing of ideas.

- Allow enough time for the student to fully express her ideas; always wait for a few seconds before proceeding to the next question.

- Review the interview transcripts several times with different investigators; this will allow multiple perspectives in interpreting the responses given by the interviewees.
7.3.3 Use of Interview for Students’ Evaluation

The following four-step model has been designed for using the interviewing process:

1. Introduce the process
   Teachers first examine responses from previous interviews and compare responses of students of differing grade and ability levels. This task provides teachers with an awareness of the diversity of children's ideas and prepares teachers for the task of analyzing the responses of their own students.

2. Provide readings
   Teachers are directed to read several articles focusing on conducting and interpreting interviews. (See Good, 1977; Kuehn & McKenzie, 1988; Osborne & Freyberg, 1985; Stepans & Kuehn, 1985). This task continues to prepare the teachers for analysis of their interviews and helps teachers recognize the importance of questioning strategies.

3. Conduct Interviews
   Teachers conduct interviews with their own students, following a script provided by the course instructor and interviewing a minimum of four students. Students are interviewed individually. It is important that teachers select students of various ages or ability levels. Student responses are audio taped so the teacher can carefully review each interview.

4. Share Results
   After the interviews are conducted, teachers share their insights orally with their classmates and in written.

7.4 PORTFOLIO ASSESSMENT

A portfolio is a purposeful collection of selective significant samples of student work accompanied by clear criteria for performance which evidence student effort, progress or achievement.

A portfolio is different from a folder in that it includes:
- Explicit guidelines for selection
- Comprehensible criteria
- Clear objectives of Selective and significant pieces
- Students’ self-reflection pieces
- Evidence of student participation in selection of content

A portfolio can exhibit the student's, progress and achievement in several areas. The potential of portfolios to assess student performance and curriculum outcomes related to attitudes and professionalism is the major driver for the use of portfolio assessment. Various assessment tools such as tutor rating scales, peer ratings and patient feedback can be included in the portfolio to provide evidence of the student’s or trainee’s performance.
The results of other examinations, however, can also be included within the portfolio framework: written tests, such as multiple-choice question (MCQ) examinations, that assess the candidate’s knowledge; other written tests, such as extended matching item (EMI) questions or short-answer questions (SAQs), that assess the candidate’s application of knowledge; and the objective structured clinical examination (OSCE), which assesses the candidate’s competence in a simulated examination setting. Thus, assessment results at all four levels of Miller’s pyramid may be included within the portfolio framework to provide a holistic view of the candidate. Portfolio assessment can be used for formative assessment, for summative assessment or both. This makes the portfolio a flexible and robust assessment method.

Pros and Cons of Portfolio Assessment

Pros
1. Provides tangible evidence of the student's knowledge, abilities and growth in meeting selected objectives which can be shared with parents, administration and others
2. Involves a considerable amount of student choice - student-centered
3. Involves an audience
4. Includes a student's explanation for the selection of products
5. Places responsibility on the students by involving them in monitoring and judging their own work
6. Encourages a link between instructional goals, objectives and class activities
7. Offers a holistic view of student learning
8. Provides a means for managing and evaluating multiple assessments for each student. The portfolio provides the necessary mechanism for housing all the information available about a student’s learning. It includes a variety of entries including test scores, projects, audio tapes, video tapes, essays, rubrics, self assessments, etc.
9. Allows students the opportunity to communicate, present and discuss their work with teachers and parents.

Cons
1. Takes time
2. Present challenges for organization and management

7.4.1 Types of Portfolio (Increasing Breath and Increasing Depth)
Several kinds of portfolio can be organized.
They are:
- Showcase (to display the students best work to parents and administrators)
• Outcome-based or assessment (to fulfill requirements and goals set by district)
• Working, process, or collection (ongoing, developmental)

7.4.2 Steps in the Portfolio Assessment Process
• Identify purpose
• Select objectives
• Think about the kinds of entries that will best match instructional outcomes
• Decide how much to include, how to organize the portfolio, where to keep it and when to access it
• Decide who selects the entries (the student, the teacher, both)
• Set the criteria for judging the work (rating scales, rubrics, checklists) and make sure students understand the criteria.
• Review the student’s progress hold portfolio conferences with students to discuss their progress

How to Implement Portfolio Assessment?
Portfolio assessment has five stages:
1. Collection of evidence of achievement of learning outcomes
2. Reflection on learning
3. Evaluation of evidence
4. Defense of evidence
5. Assessment decision

1. Collection of evidence of achievement of learning outcomes
The student collects evidence of achievement of the learning outcomes during his or her day-to-day learning activities, interaction with patients or other studies. The evidence can be anything from a tutor rating to evidence of visiting a patient’s home, which the student thinks has helped her or him achieve the curriculum learning outcome(s). “The evidence in portfolios,” suggest Friedman Ben-David et al., “is limited only by the degree of the designer’s creativity. ”Some categories of information that can be included in the portfolio are best essays, written reports of research projects, evaluations of performance (e.g., tutor reports, checklists, patient evaluations), videotapes of interactions with patients or peers, records of practical procedures mastered, annotated anonymized patient records and curriculum vitae. Much of this material will be submitted by the student for marking and feedback during the portfolio-building process.
Traditionally, most of the evidence collected has been paper-based. Portfolios, however, are increasingly becoming computer-based (i.e., e-portfolio). E-portfolios have addressed, at least partly, concerns regarding the volume and portability of the traditional paper-based portfolio. Although students may collect any evidence they wish, this tends to make the portfolio unmanageable—a drawer for “everything but the kitchen sink.” A broad structure for the portfolio is needed to standardize content for summative assessment purposes. Balance is required between structure, to provide suitable material for summative assessment and student selection of portfolio content to express the individuality and creativity of the candidate. “It is advisable to add structure to the assessment but to refrain from over structuring, as this tends to trivialize the measurement.” The student will have to exercise discretion not only regarding the type of material to be included in the portfolio but also in deciding the volume of portfolio material. “Too much information can create an unwieldy collection of documents that only the owner can decipher,” while too little will be an underrepresentation of achievement. Creating an assessment blueprint, a grid that engages curriculum outcomes with curriculum content, is one of the best ways to ensure that the portfolio has sampled all the content and represented all the outcomes in appropriate amounts and proportions. The portfolios used to revalidate Tayside general practitioners provide an example of a framework that has achieved the required balance between outcomes and content without compromising either. Summarizing portfolio content, Stecher states that portfolio content should have diverse products of student learning; be cumulative (i.e., should contain work completed over a period of weeks or months); and be embedded in instruction (i.e., entries are drawn from ongoing work).

2. Reflection on Learning

Reflection, according to a model developed by Schon as quoted by Challis, is “revisiting an experience after the event, in order to extract the principles and effectively ‘bank’ these for future use.” This is “reflection on action.” The reflective process should be directed to promote learning, personal and professional development, and improvement of practice. In the context of portfolio assessment, reflection must answer four questions relating to a learning experience: What did I learn? What do I still need to learn? What
resources did I use for further learning? and What further learning was achieved?

3. **Evaluation of Evidence**

   Once the student submits the portfolio, assessors will evaluate the quality of the evidence it contains. The assessors rate the student’s achievement of the learning outcomes on rating scales, anchored with precise and specific descriptors of behavior at each point on the scale. Such rating scale rubrics help benchmark the passing standard in the standard-setting process of portfolio assessment. Because of the subjective nature of the assessment decisions involved in the use of rating scales, ratings of several examiners must be collated to arrive at a reliable evaluation of the portfolio evidence of a particular student. Evaluation of evidence provides feedback to both assessor and assessee. The assessor, by analyzing the evidence of performance in the portfolio, finds out how successful the teaching/training has been and what changes are needed for improvement. In practice, it is crucial that the evaluation of the evidence be an ongoing process, with feedback provided for students throughout the period of portfolio building. The feedback indicates to the assessee what her or his strengths and weaknesses are and what areas need improvement. For this reason, it is imperative that students know how to interpret the assessors’ ratings. Involving students in designing the rating scales, as reported by Williams in a middle-school portfolio assessment in the United States, may be an effective way of sharing the assessment criteria with the students. Profiling students’ progress toward the learning outcomes over time will facilitate this feedback process.

4. **Defense of Evidence**

   At this stage, the examiners interview the candidate to probe how well the portfolio has reflected his or her achievement of the learning outcomes. The examiners use the interview to confirm or refute the decisions they made regarding the candidate’s strengths and weaknesses in terms of the learning outcomes when they read and evaluated the portfolio in stage 3. It is our experience at Dundee Medical School that if the examiners, following their initial evaluation of portfolio evidence, think that a student has clearly passed, subsequent discussion of portfolio evidence between the student and
the examiners may not be essential. The stage of defending portfolio evidence might be restricted only to borderline, failing, and honors candidates. However, the students appreciate the opportunity to discuss their portfolio work with the senior faculty, and so this portfolio assessment stage has been retained.

5. **Assessment Decision**
Pre-validated rating scales with clearly defined descriptors are used to assess the evidence. The expected standard of performance, as a point on the rating scale, is agreed by the examiners prior to the assessment. Any disagreement regarding the standard of student performance is resolved by discussion between the examiners after the student interview. In essence, the assessment decision is taken by consensus among the examiners.

7.5 **COMPUTER-ASSISTED TESTING AND FUTURE TRENDS IN ASSESSMENT**

It refers to the use of computers to assess students’ progress. The assessments can vary in format: either consisting of a pre-printed paper test on to which students mark their responses, which are then processed automatically using an optical mark reader; or involving the direct input of students’ responses into a computer terminal. Computer-based assessments may be stand alone and specific to certain machines within a computer lab; based on a local network (intranet); or, as is increasingly common, web based. The nature of the assessments may also differ. They can be diagnostic – to determine students’ knowledge prior to starting a course, potentially allowing amendments to their specific course design. They can be formative and include ongoing feedback either during the assessment or after. They may be ‘scored formative’, allowing ongoing comparison of a student’s progress over a period of time, possibly replacing an end-of-term summative assessment.

Alternatively, they may be summative, contributing to a student’s end-of-year mark. Depending on circumstances, such tests can be either supervised or non-supervised, with the option of allowing students to check their own progress through self-assessment. Although more commonly used for testing lower-order skills (such as knowledge, understanding and application), when properly formulated they can also be used for testing higher-order skills (analysis, synthesis and evaluation). Their nature allows the automation of what was previously a very time-consuming task: that is, marking and monitoring progress.
Like other tools, a properly constructed CAA package can be very efficient in achieving its designated tasks.

Computers can be used to deliver, mark and analyse student assignments or examinations. CAA can be delivered using stand-alone computers or online, using an Intranet or Internet. There are two main forms of CAA. The first is where students input their answers via the computer. The second is where students input their answers onto a pre-printed paper test. This is then fed into an Optical Mark Reader (OMR), which rapidly processes the paper forms by scanning for marks or crosses. CAA tests can be supervised or non-supervised, formative or summative. There is also a diagnostic role for CAA in determining students’ prior knowledge of a subject. The growing interest in online learning within tertiary education is leading to a recognition that any plans to implement online learning should also include a consideration of CAA.

A number of reasons may be given to justify the use of assessments and need to be borne in mind when considering the appropriateness of CAA tests. They may include:

- Helping the student learning process through formative assessments.
- Grading students’ work through summative assessments.
- Being able to regularly evaluate student progress.
- Helping to monitor how effective current teaching and learning strategies may be.

**Pedagogical Advantages of CAA**

- It enables the assessment of a wide range of topics very quickly, with an associated reduction in the time that lecturers dedicate to marking.
- The need for double marking is totally eliminated. This time and resource saving allows more regular assessment than might otherwise have been possible, consequently enabling more detailed knowledge of students’ progress and quicker identification of problems.
- Tests can be tailored to match students’ abilities and, with formative assessments, it may be possible for content to be varied automatically as the test itself progresses, matching students’ weaknesses as they emerge during the test (‘computer adaptive testing’).

**Administrative Advantages of CAA**

- The saving of time in supervision, invigilation and marking, and a reduction in subjectivity and human error in the marking process itself.
- When dealing with large groups of students, the time and resource saving can be of a significant order. Given the computer-based nature of the
exercise, substantial reductions in printing costs should be achieved when assessments are updated or altered.

- Statistical evaluations of the results can be generated automatically, with no additional keying in of marks, thus reducing evaluation time.
- Successful integration into, and co-ordination with, student records and university information and management systems.

It is apparent that education policy is increasingly conferring a central strategic role to evaluation and assessment as indispensable tools for improvement, accountability, educational planning and policy development. In the last two decades, most countries have introduced a wide range of measures intended to improve evaluation and assessment at all levels from the student to the school system itself. These have done much to stimulate public awareness of evaluation and assessment and to develop an evaluation culture within school systems.

The move towards an increase in the use of technology – for teaching, learning and assessment – has resulted in changes in pedagogy in mainstream education away from transmission approaches towards other models such as those of Flipped Learning, Deep Learning and 21st Century skills. These new pedagogies emphasise collaborative projects, problem solving, and active learning, all of which can be supported by digital technology. Such models are also based on principles of constructive alignment between learning activities and tasks, learning outcomes and assessment tasks; again there is some evidence that technology can enhance such constructive alignment. Many language teachers are already familiar with collaboration, projects and active learning as communicative tasks and activities often incorporate such principles already; however, they may be less familiar with how to use digital technology to promote such pedagogy.

Secondly, an ever greater focus on technology for creating digital projects such as slideshares, videos and collaborative documents raises concerns of digital literacy such as copyright and plagiarism, the use of Creative Commons for incorporating others’ work into web authoring, and issues of privacy and academic integrity. These, too, are new aspects of assessment in second and foreign language education that will increasingly need to be taken into account.

All these changes in technology, pedagogy, and assessment practices mean that teacher roles will change; or at least will be under some pressure to change. In a provocative paper, Godwin-Jones (2015) identifies a number of ways in which he recommends that future language teachers could adapt to technology.
These are:
1) to learn to code in order to be able to evaluate and adapt new software;
2) to learn how to include mobile applications in lessons;
3) to learn how to use technology in context and be able to assess its applicability to one’s own situation; and,
4) to use technology to participate in global learning opportunities.

In brief, Godwin-Jones emphasises that teachers need to learn ‘technology fluency’. They can do this in various ways such as by setting up their own web domain for the dissemination of e-Portfolios and the adaptation of a ‘maker culture’ in which participants create and share various kinds of products. Godwin-Jones sets a high bar with his ideal teacher being technologically literate, actively involved in researching and experimenting with new approaches, and committed to cross-cultural understanding; however, it is a useful set of aspirational goals to consider for future teacher development. In contrast, it is important to point out that there are a number of critical voices that urge educators to be vigilant in the apparent rush towards the ever-increasing use of technology in education.

Selwyn (2014) is particularly prominent in advising teachers and institutions to be ‘purposively pessimistic rather than unrealistically optimistic’ in their approach to digital technology and to question whether new digital systems really do empower students and teachers or are just mechanisms of control and surveillance. Selwyn argues very persuasively that digital technology may offer only limited pedagogic value to staff and that students vary enormously in their ability to make best use of its affordances. He concludes with a number of suggested ways to ‘weak’ our use of digital technology to overcome these objections including the need to challenge the way language is used with regard to technology so that it does not obscure what it can really achieve.

7.6 SELF ASSESSMENT QUESTIONS

1. Explain oral questioning as an assessment tool. Also describe purpose of questioning.
2. Describe Guidelines for questioning in classroom.
3. What do you understand by Peer appraisal? Explain different techniques applied in peer appraisal.
4. Write brief note on:
Guess who’ techniques
Socio-metric techniques
5. Define an interview? Explain different types of interviews.
6. Discuss strengths and weaknesses of interviews.
7. Justify the role of interviews for students’ evaluation.
8. Explain portfolio assessment and its types.
9. Explain different steps involved in the portfolio assessment process.
10. Elaborate computer-assisted testing and future trends in assessment.
INTERNET SOURCES

- http://public.callutheran.edu/~mccamb/questioning.htm
- https://www.qub.ac.uk/directorates/AcademicStudentAffairs/CentreforEducationalDevelopment/FilestoreDONOTDELETE/Filetoupload,213702,en.pdf
- https://www.economicsnetwork.ac.uk/handbook/printable/caa_v5.pdf
- https://www.alt.ac.uk/sites/default/files/assets_editor_uploads/documents/eln004.pdf
UNIT - 8

ASSEMBLING, ADMINISTERING AND APPRAISING TEST

Written By: Shahida Parveen
Reviewed By: Dr. Muhammad Tanveer Afzal
CONTENTS

Introduction..................................................................................................................................................161
Objectives .......................................................................................................................................................162
8.1 Assembling the Classroom Test ...................................................................................................................163
  8.1.1 Alignment of the Items with Table of Specification .................................................................................164
  8.1.2 Different Arrangements and Placement of the Items ...............................................................................166
  8.1.3 Assembling Different Versions of the Tests by Replacing the Items and Their Position .........................167
8.2 Administering and Scoring ..........................................................................................................................168
  8.2.1 Development of Instruction for the Administration of Test ......................................................................168
  8.2.2 Role of Invigilators ..................................................................................................................................169
  8.2.3 Scoring of Objective Type Tests (through key) .........................................................................................170
  8.2.4 Scoring of Subjective Type Tests (through rubric) .................................................................................171
8.3 Appraising Classroom Test ..........................................................................................................................172
  8.3.1 Reviewing Test Items .............................................................................................................................173
    8.3.1.1 Item Analysis .....................................................................................................................................174
  8.3.2 Standardized Testing .............................................................................................................................174
  8.3.3 Classical Test Theory ...........................................................................................................................175
  8.3.4 Item Response Theory ..........................................................................................................................177
8.4 Self Assessment Questions ...........................................................................................................................179
Internet Sources ..............................................................................................................................................180
INTRODUCTION

A prospective teacher should know the all requirements of teaching learning activities in which assessment is a continuous part of formal learning process to evaluate students’ learning outcomes and competencies. The purpose of the assessment process is to assess the competencies of individuals to practice selected functions effectively and efficiently (Marry, 2015). Test is used as basic tool of assessment process to evaluate the achieving of learning objectives. The development of test items, assembling of developed test items in different formats and arrangements is a challenging task regardless which type of items required in a test. Test item development can proceed only when a clearly agreed upon set of objectives is available (Cohen & Wollack, 2013). Assembling of test is myth of test development process, any mistake at this stage leads to fail in achieving the objectives of test. Assembling of test has been required three more steps in process of during or post writing test items;

- Alignment of test items with Table of Specifications (TOS).
- Different arrangements and placement of the items.
- Assembling different versions of the tests by replacing the items and their position.

After assembling a test next stage is administration of test. Administration of test is group activity involving administrator of assessment process, Head invigilator and two or more invigilators depending upon number of students enrolled for test. Test directions should be designed to maximize understanding of the task being presented and to minimize confusion on the part of test-takers as to what they are expected to do. According to Gronlund (2003) selecting the type of test item is a challenging task to justify the requirements of assessment process. So the type of test items should be selected on the basis of learned content, learning objectives and desired cognitive domain. The scoring of test items should be followed the prescribed rules of scoring. Test developer is required professional improvement with the passage of time due to fastly changing curriculum plans and learning objectives of instructional process. Accountability and credibility of testing and assessment process should be enhanced decreasing the interaction of students during test conduction. Different arrangements and placement of the items are used to minimize the students’ interaction during test conduction.

Appraising the test means whether it is used as classroom test or as estimating the worth of learning process or as value of assessment process to set according the required standards and norms of test, learning process or assessment. In order to maintain the currency and relevancy of the assessment, item development is an ongoing process (Marry, 2015). With regard to the presentation of test materials,
assessment developers should take into account formatting considerations (e.g., fonts, font sizes, and the location of line breaks in sentences and paragraphs). Appraising the test is required further steps to be taken such as;

- Reviewing the test items
- Item analysis (item difficulty & item discrimination)
- Standardized Testing
- Test theory (classical test theory & item response theory)

**OBJECTIVES**

After study this unit, the prospective teachers will be able to:

1. assemble the classroom test by using different formats and arrangements.
2. develop guidelines to administer the test and administer it accordingly.
3. perform the role of invigilators.
4. score the objective type test and subjective type test.
5. review the test items qualitatively as well as quantitatively.
6. develop and conduct the standardized test.
7. describe critically “Item Response Theory (IRT)” & “Classical Test Theory (CTT)”.
8.1 ASSEMBLING THE CLASSROOM TEST

Assembling of test is myth of test development process, any mistake at this stage leads to fail in achieving the objectives of test. Test is a tool to assess the students’ learning abilities and skills regardless it is used in formative assessment or summative assessment. Its main purpose is to identify the students’ cognitive developmental process and learning output. A test is used to ensure the achievement of learning objectives during or post instructional process. Which type of test items should be included? What should be given time limit to attempt the test? What should be weigh age of all type of test items? Which level of cognition to be measured? These questions are raised before writing test items and answered through Table of Specification (TOS). The selection of item types and test format should be based on the kinds of skills to be measured and not on some personal like or dislike for a particular item format. (Cohen & Wollack, 2013)

Development of Table of Specification (TOS) is first step in constructing a test rather it used during instructional process or end of the instructional process. It is based on learning objectives of the instructional process and learning objectives of the instructional process are expectations of curriculum and policymakers. These are presented in written form to the teachers. They can plan their instructional strategies in the light of these learning expectations. These expectations should be shared with learners so that they can be more motivated towards learning process. Textbooks are major source of contents to be shared and learned in schools. Now the learning objectives are written in every textbook before starting each and every unit. But how is it possible to judge whether these learning objectives are achieved or not?

Written Test is used to assess the learning objectives of any content are achieved or not in a formal assessment process. The numbers of test items, the type of test items, the weigh age of items and time limit of test items to be attempted, are planned before writing test items to avoid any ambiguity. This plan is called Table of Specification (TOS). A Table of Specification (TOS) can be used to help teachers frame the decision making process of test construction and improve the validity of teachers’ evaluations based on tests constructed for classroom use. (Fives, Nicole, & Barnes, 2013). It is useful for teachers to spend appropriate time in teaching to meet learning objectives and create a linkage between assessment process and learning activities. The TOS is useful to organize the planning process of assessment.

Item writing is a creative work required expert based knowledge about contents, skills and traits to be tested. Item writing is an arduous task requiring not only
mastery of the subject matter, but also an understanding of the examination population and mastery of verbal communication skills (Marry, 2015). Item writer should well aware about age group of students and learning objectives to be assessed. He/she should have knowledge about assembling of test according to already developed rules of particular test. Assembling of test has been required three more steps in process of during or post writing test items;

- Alignment of test items with Table of specification (TOS)
- Different arrangements and placement of the items
- Assembling different versions of the tests by replacing the items and their position

These thrice steps are mostly considered essential in assembling of any test and their detail discussion are given below.

**8.1.1. Alignment of the Items with Table of Specification:**

The term “alignment” has been used in education for many decades of previous century. Alignment is worked as linkage between assessment process and instructional strategies. But it should not be compromised on essential components of assessment; validity and reliability. “Alignment is the degree to which the components of an education system, such as standards, curricula assessments and instruction work together to achieve desired goals” (Khalid, Azeem, & Gondal, 2011). Alignment of learning objectives with assessment process plays an important role in knowing about what level of students’ learning has been achieved.

Webb (1997) says,

“Students are being assessed on what they are expected to know”

“TOS is referred to as test blue print that helps teachers to align objectives, instruction and assessment” (GronLund & Linn, 2000; Fives, Nicole, & Barnes, 2013; Wollack, 2014). It is a challenging task to align the test items with TOS but careful analysis and expert review can make it possible to improve the quality of test. Test developer is required professional improvement with the passage of time due to fastly changing curriculum plans and learning objectives of instructional process. TOS is essential source of aligning the test items with learning objectives of content. The TOS provides a framework for organizing information about the instructional activities experienced by the student (Fives, Nicole, & Barnes, 2013).

To align the items with TOS is required some characteristics among item writers. An item writer should have expert level knowledge about the material being
tested to well align the items with TOS. He/she should have expressive writing skill to write items in simple and appropriate language. This task is needed great mastery in test development field. The goal of test development is to link the steps as closely as possible, newly written items should specifically map to the outline of the test. (Mihari, 2010). The teachers can use TOS to improve their professional judgment when creating items of test and aligning them to learning objectives.

Webb (1997) has presented three approaches to align the test items with TOS, which are still considered recommendable by researchers.

- Sequential Development
- Expert Review
- Document Analysis

These approaches are discussed below;

- **Sequential Development:**
  The alignment of test items with TOS can be considered as sequential process. “Developing standards, framework, policy and assessment in sequence has the advantage of proceeding in a logical process and after the development of the first document, having known criteria for the development of subsequent documents”. (Webb, 1997) But here a question is raised, what is known criteria for alignment and development of subsequent document? The known criteria are educational standards, educational policy, curriculum plan and learning objectives of the contents. The TOS is used to define criteria of particular test. In defining criteria these steps should be followed;
  - Individual differences among students should be kept in mind before defining criteria
  - Equity and fairness should be attributes of criteria.
  - Pedagogical implications for criteria should be possible.
  - Realistic and adaptive system of testing should be added.

Prospective teachers should follow these steps to avoid any ambiguity and biasness in defining criteria. They should assess each and every type of students equally and fairly applying different pedagogical implications and adaptive system of testing to fulfill the requirements of teaching learning activities without any discrimination.

- **Expert Review:**
  Content area specialists are necessary to review any assessment procedure specially to confirm the alignment between learning objectives and test
items. Experts can be hired to check the alignment level of test items with TOS. Different organizations of assessment have recruited content specialists for expert review of test items to align more and more with TOS and learning expectations.

- **Document Analysis:**
  The textbooks are source of material being to be taught and assessed during learning process and post learning process. These are assumed as expression of educational policy and curriculum plan of a state. These are used for Document analysis in alignment. Alignment can be judged by coding and analyzing the documents that convey expectations and assessment (Webb, 1997).

8.1.2 Different Arrangements and Placement of the Items

Test plays an important role in determining learning achievement and certifying attainment of students during assessment. Accountability and credibility of testing and assessment process should be enhanced decreasing the interaction of students during test conduction. If all students in examination hall have been responded similar test then the element of cheating may be involved and credibility of test scores may be affected. Different arrangements and placement of the items are used to minimize the students’ interaction during test conduction. These arrangements and placement of items should have similar difficulty level. Difficulty level of all arrangement should be similar to make trustable and worthwhile classroom test. Test Items should be of appropriate difficulty and commonly three types of arrangements are used according to difficulty level;

- Easy to hard
- Hard to easy
- Random distribution

Many researchers (Macnicol, 1956; Anstasi, 1976; Sheppard, 1994; Gerow, 1980; Allisons, 1984; Soyemi, 1980; Perlini, Lind & Zumbo, 1998) studied and researched in arrangements of test, and concluded that the arrangements of items were affected the performance of students and mostly liked the arrangement of items “Easy to hard”. While “Hard to easy” or “Random distribution” had been created more test anxiety among candidates. The impact of item order influences the performances of students in MCQS (Estey, 2010).

Gronlund (2003) describes about the arrangements of test that in assessment of student’s academic achievement, the arrangement of test items may influence the performance of student, so some precautions should be taken such as;
- To group items together that measure same learning ability.
- All items of same type should be grouped together.
- The items should be arranged in terms of increasing difficulty easy to hard.

### 8.1.3. Assembling Different Versions of the Tests by Replacing the Items and their Position

Assessment process is increased its accountability and objectivity in nature through assembling different versions of the tests by replacing the items and their position. Fair assessment process may be conducted to limit the students’ interaction or any external illegal support. Arranging test items in different versions is a complex task and required expert level knowledge of standardized testing. Computer Adaptive testing should be used to assemble different versions of test. Standardized testing organizations in all over the world have been becoming active users of computer Adaptive testing. But Reliability and validity of test should not be compromised during assembling different versions of test. Appropriate difficulty level should be maintained in all versions of test.

The way in which an assessment administered to students is especially important for standardized assessments (Zucker, Galindo, Grainger, & Severance, 2004). In testing situations, the use of alternate test forms constructed with the same items presented in different order is one of the strategies for deterring copying and enhancing test security in test administrations (Chiu, 2012). Commonly five or six versions of test are used in one test and use of item banking may be effective to select items for different versions of test. An item bank is a collection of test items that may be easily accessed for use in preparing examinations. The difficulty level of items should be checked continuously and regularly. Item banking can be a useful way for educational systems to monitor educational achievement (Stanley, 2009).

Developing an item bank requires understanding the construct of interest, writing items at multiple levels of difficulty and then validating the items to make sure that they represent the relevant construct. Items should be developed and selected based on appropriate difficulty level and discrimination level. Some items may require to be removed from the item bank due to vastly changing the curriculum plans and new items should be incorporated into the bank on regular basis (Stanley, 2009).

<table>
<thead>
<tr>
<th>Activity 8.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Develop a table of specification for the construction of test items from unit 1 of the 9th class English textbook. Formulate MCQs, short questions and two essay type items aligning them with Table of specification. Prepare five different sets of test by using different arrangements.</td>
</tr>
</tbody>
</table>
8.2 ADMINISTERING AND SCORING

After developing a test next stage is administration of test. Administration of test is group activity involving administrator of assessment process, Head invigilator and two or more invigilators depending upon No. of students enrolled for test. Administrator is responsible for developing instructional plan for test, while invigilators are responsible for peaceful conduction of test. The detail discussion is given in next.

Scoring is next stage of test soon after its administration and conduction. There are commonly two types of test items assessed; objective type and Essay type. The scoring of test items is complicated task requiring expert level knowledge of contents. The objective type items are scored with Key and subjective type items are scored with the help of Rubrics. The detail of scoring is given in next.

8.2.1 Development of Instruction for the Administration of Test

In standardized testing the procedure of administration is very affected on the scores of students. To avoid any ambiguity the development of instruction for the administration of test is essential part of testing. During the development process, standardization establishes uniform procedures for using the assessment so that observation, administration, equipment, materials and scoring rules will be the same for every student to whom it is administered (Millman & Greene, 1993; Zucker, Galindo, Grainger, & Severance, 2004). The guidelines for administration the test should be clear, precise and easy to follow. Mostly these guidelines are provided in written form (booklet) to the test organizers in real test settings. These test organizers are assigned different roles according to test requirements. The physical conditions of the examination hall should be carefully monitored and guided the each and every student to understand the instructions about the testing conditions. Disable students should be facilitated according to their needs. Testing environment should be comfort for all students. Test instructions should be written in such format which is easy to read for everyone. The goal of this uniformity is to make the assessment’s results as objective as possible so that they can be considered valid and meaningful when used to compare the assessed qualities of students (Zucker, Galindo, Grainger & Severance, 2004).

Test directions should be designed to maximize understanding of the task being presented and to minimize confusion on the part of test-takers as to what they are expected to do. Administration procedures that protect the security of the test help to maintain the meaning and integrity of the score scale for all examinees.
8.2.2 Role of Invigilators

There are two types of invigilators working in examination hall; head invigilator and invigilator. The role of invigilator is required some attributes whether he/she working as invigilator or as head invigilator. These attributes are two types; Personal attributes and Professional attributes.

**Personal Attributes**
- He/she should possess strong communicational skills (speaking and writing).
- Test conduction is a group activity and coordination of all members of group is essential part of this task, so positive attitude and ability to cope with others are required among invigilators.
- He/she should work according to rules of administration and not show any misconduct.
- He/she should flexible in working and self controlled personality.
- He/she should have patience to deal with aggressive students in examination hall.
- He/she should be an agent of providing safe and relaxed environment in examination hall.
- He/she should have self motivation and visionary approach to perform his/her duties and avoid to work as boring personality.

**Professional Attributes:**
- He/she should be qualified academically and professionally to compete the prescribed standards of invigilation of examination system.
- He/she should possess organizational skills to organize the test conduction activities.
- He/she should able to maintain record management updated.
- He/she should expert in time management to avoid any disturbance and distraction in examination hall.
- He/she should handle carefully any violation of rules and regulation during conduction of exams.
- He/she should follow the instructional rules of test; for example distribution of test according to version wise or code wise and distribution of test according to time limit of test.

**Responsibilities of Head Invigilator and Invigilators**
They should be trained properly to maintain all environmental factors which may be caused hindrance in conduction of the test. They have to fulfill some responsibilities in any type of test venue such as:
- They have to maintain and secure the stationery, examination papers and other material related to exams according to prescribed arrangements.
They guide the students to sit in examination hall according to seating plan which has been already planned according to No. of enrolled students.

They distribute question papers and answersheets carefully following the written rules of test conduction.

They invigilate during exams and solve the issues raised by students following the rules and regulations.

They deal strictly any irregularity created in examination hall.

They maintain attendance record of students according to their allocated Registration No. and code No. of Question papers.

They make arrangements in provision of peaceful physical environment to students involving; provision of proper light, drinking water, proper cooling or heating depending on weather condition and proper seating arrangements.

At the end of test, they collect the answer sheets carefully and arranging them orderly.

They pack the answersheets and remaining unsolved question papers accordingly to send the test administration authority.

They make seating plans for next test which may be held second time of the same day or next working day.

8.2.3 Scoring of Objective Type Test (through Key)

According to Gronlund 2003, selecting the type of test item is a challenging task to justify the requirements of assessment. So the type of test items should be selected on the bases of learned content, learning objectives and desired cognitive domain. There are basically two types of test items; objective type test and essay type test. Objective type test is further categorized as; MCQs, completion test, Match the columns and binary options (True, False or Yes, No). The usage of MCQs are becoming more popular in test but it has becoming a great debate that MCQS cannot measure high order thinking, but with appropriate efforts in developing MCQS can be resulted to measure high order thinking. Commonly MCQs are used for multipurpose; to identify some pattern, to cover more content, measuring high order thinking and logical reasoning to some extent depending on the context of testing. These are considered more reliable and valid. They should be developed to avoid guessing item writing with the help of TOS (Croft, Guffy, & Vitale, 2015). The checklist developer should assure that the checkpoints in each category are sequenced logically and functionally (Stufflebeam, 2000).
A rater is a person who scores the responses based on scoring keys. Raters are typically trained and certified, and also retrained after a certain period of certification, by a testing organization. After being certified, a rater scores test-taker responses based on scoring guides. The scoring keys for the assessment should be the focus of the training and retraining of raters. (Ockey, Yong, & So, 2004). Scoring of objective type test is easier than scoring of subjective type test. Its scoring is considered more objective and unbiased as compare to subjective type test. The scorer’s attitude, emotions or mood cannot influence the scoring of objective type test. The key for scoring is already prepared to avoid any ambiguity in scoring. The experts of contents should be prepared the key. The item writer of objective type test may write the key for scoring. The objective type test may be consisted of different versions or having different arrangements of items. The scorer should carefully observe the exact code of version and then start the marking. He/ she should highlight the answers on key with visible ink to avoid visual fatigue. After marking the test he/she should review it again to make sure about scoring. His/her careful review of scoring may be affected the scores of students.

8.2.4 Scoring of Subjective Type Tests (through rubric)

Developing of subjective type test is commonly considered easy as compare to its scoring. Its scoring is required more time and efforts. High order thinking is measured through subjective type test. Rubrics should be prepared to score subjective type test. Without rubrics a scorer’s personal interest, emotions and stereotypes are affected the scoring of subjective type test. Then the results of test and objectivity of scoring are biased. To avoid any ambiguity in scoring, rubrics should be developed before scoring. In developing rubrics as scoring key for objective type test items, one should well aware the objectives of test, the ability levels of the students, the requirement of assessment and weigh age of different portions of required subjective type response. Rubrics are made to minimize subjective nature of scoring. Rubrics are evaluation checklist that clarifies the criteria at least should be considered when evaluating something in a particular
area; aids the evaluator not to forget important criteria; and enhances the assessment's objectivity, credibility and reproducibility (Stufflebeam, 2000).

A scorer should keep in mind all rules of scoring which are prescribed in Rubrics. Scoring of subjective type test is difficult to grade. Grading is often affected by the verbal fluency in the answer, handwriting, presence or lack of spelling errors, grammar used and the subjective judgments of the grader (Cohen & Wollack, 2013). So to make sure the objectivity of scoring the following guidelines are helpful for subjective type test items in a consistent and meaningful way.

- Construct a model answer for each item and award a point for each essential element of the model answer.
- No effect of good hand writing and fluency of language except the language test.
- Graders should not be aware directly to students; they may be allotted secret code numbers as practiced in BISE Exams of Punjab.
- Same criteria for scoring to all students.
- Two graders should be graded same items to compare scoring.
- Head grader should be randomly assessed the scoring of graders.
- Appropriate training for grading should be provided to graders.

<table>
<thead>
<tr>
<th>Activity 8.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>This activity is consisted of two parts:</td>
</tr>
<tr>
<td>a. Administer the test which had been developed in Activity # 8.2, on the students of Grade 9 of any local school with the coordination of English teacher, write the instructional manual and perform the role of invigilator fulfilling all requirements of test conduction.</td>
</tr>
<tr>
<td>b. Allotting the secret roll # to every answer sheet and then scoring the test accordingly.</td>
</tr>
</tbody>
</table>

8.3 APPRAISING CLASSROOM TEST

The entire test development process contributes to construct validation, whatever the nature of the test (Anastasi, 1973). Appraising the test means whether it is used as classroom test or as estimating the worth of learning process or as value of assessment process to set according the required standards and norms of test, learning process or assessment. In order to maintain the currency and relevancy of the assessment, item development is an ongoing process. With regard to the presentation of test materials, assessment developers should take into account formatting considerations (e.g., fonts, font sizes, and the location of line breaks in sentences and paragraphs. Appraising the test is required further steps to be taken such as;
Reviewing the test items
Item analysis (item difficulty & item discrimination)
Standardized Testing
Test theory (classical test theory & item response theory)

Here is given brief description one by one.

8.3.1 Reviewing Test Items

“It is relaxation time” when task of writing test items is completed. But stop; it is not ending; it is a starting point to a long journey of test reviewing to ensure validity and reliability in a context of quality controlling authorities. Irrelevant information can lead to possible challenges by examinees that the item is not valid, namely not focusing on measuring what is supposed to be assessed (Corporation, 2011). Each test is constructed through the coordinated efforts of the appropriately qualified item writers and examination developers who are subject matter experts, and psychometric experts who provide information on the measurement characteristics of the items and/or tests (Marry, 2015).

A test item reviewer should able to;
- Understand students’ level of learning.
- Know about the learning objectives of unit or content.
- Aware about the curriculum standards about particular course.
- Learn about testing techniques.
- Know about test development process.
- Use of technology.

Careful reviewing of test items should not be compromised on test validity, test reliability, ease to administer and ease to understand (Gronlund, 2006). The aim of the reviewing the items is to differentiate the items which are not measuring the learning objectives according to expectations of test. After items are presented on a test, they are analyzed statistically to identify difficulty level of items and discriminating power of items. The statistical analysis provides clues for the subject matter experts with regard to how well the content of the item yielded useful information about candidate ability (Marry, 2015). Statistical analyses are informative with regard to how well each item on an assessment is functioning. Items should be analyzed with regard to their difficulty level (how hard the item is), discrimination value (how well the item separates high- and low-ability test-takers), and association (correlation) with other items, as well as test section score and total score. (Ockey, Yong, & So, 2004). The item analysis should be part of item reviewing process to examine the items one by one.
8.3.1.1 Item Analysis

Items can be examined qualitatively on the basis of their content and forms, and quantitatively on the basis of their statistical properties. Quantitatively analysis is based on Item difficulty and item discrimination of test (Anastasi, 1976). The statistical analysis provides clues for the subject matter experts with regard to how well the content of the item yielded useful information about candidate ability (Marry, 2015).

Item analysis can provide important diagnostic information on what examinees have learned and what they have not learned. There are many different procedures for determining item analysis. The procedure employed in evaluating an item's effectiveness depends to some extent on the researcher's preference and on the purpose of the test. Item analysis of a test comes after the first draft of a test has been constructed, administered on a sample and scored out. In the process of item review, the item statistics represent the performance of the item and provide guidance to the examination reviewers when revising items. Both previously tested and new items may require revision (Marry, 2015).

<table>
<thead>
<tr>
<th>Activity 8.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Take any test of Grade 10 which has been already conducted through any external assessment organization and review it carefully such as; whether it is measuring the learning outcomes or not, whether it is according to the rules or standards of curriculum or not, whether its language is appropriate or not and whether it is according to learning abilities of student or not.</td>
</tr>
</tbody>
</table>

8.3.2. Standardized Testing:

Standardized testing holds teachers and schools accountable. It is normally developed by experts of the specific field. For example in psychology, different standardized tests are used to measure social skills, attitudes or developmental phases of human beings and these tests are prepared after a long time research such as MMPI, IQ test, TAT and so on. Standardized tests are developed to have consistency in test items, administrative plan, and scoring guides. The term “test standardization” establishes uniform procedures for using the assessment so that observation, administration, equipment, materials and scoring rules will be the same for every student to whom it is administered (Millman & Greene, 1993; Zucker, Galindo, Grainger, & Severance, 2004).

The history of standardized testing is beyond the research of centuries. Anstasi (1973) was first president of American Psychological Association (APA). She researched about test, test anxiety and difficulty order of test items.
According to her,

“Standardized Tests are constructed to compare norms and criterion referenced test, the criterion generally refers to some independent, external measure that the test is designed to predict, such as direct measure of job performance used in evaluating an applicant selection test”

Testing process should be unbiased and objective by nature. Reliability is the extent to which the test would produce consistent results if it is given again under the same circumstances. (Junker, 2012). Standardization attempts to control external factors to the greatest degree possible so that the assessment is a valid measurement tool that produces meaningful results (Zucker, Galindo, Grainger, & Severance, 2004). Three factors of assessment are essential components of standardized testing;

- Validity (content, criterion & construct)
- Reliability (test re test, split half & alternative forms)
- Utility

Teachers can share the results of standardized testing. Students can be compared across the world on the basis of standardized testing. The standardization process is conducted under highly controlled conditions, including the time limits (if specified) for each test in the assessment’s battery, the materials the students may use during the assessment (such as scratch paper or calculators), and the directions for administering (Zucker, Galindo, Grainger, & Severance, 2004). Appropriate standardized procedures improve measurement by increasing consistency and test security. Consistency means that the standardized tests should be developed to administer under consistent procedures and the test-taking experience is as similar as possible across test takers. Test security consists of ways adopted to prevent the use of unfair means such as cheating and exposing the test items and content before the conduction of test. The main purpose of test construction is to develop test items of desired quality, no matter what type of test is used.

According to Junker (2012),

“A standardized test is known famously for two mantras; higher reliability is always better, a test cannot be valid unless it is reliable”.

8.3.3 Classical Test Theory (CTT)

The history of classical test theory is consisting of almost more than 150 years, when any human attribute was measured; the chance of error prevailed and
minimized the reliability and validity of measuring instrument. CTT was used over the majority of 20th century and is still used in test development (Güler, Teker, & Uyanık, 2013). Charles Spearman (1904) firstly tried to measure the error of test having an insight that there were found errors in true test scores, and these errors were randomly varied, and that they could be correlated and indexed, so he is considered as founder of Classical test theory. Gulliksen (1950) attributes the basic form of CTT, Novick (1966) made CTT safe for modern readers, and a detailed modern treatment is given in Lord and Novick (1968) (Junker, 2012). CTT is used to provide theoretical framework for analyzing, developing and recognizing the reliability of standardized tests and assessment (Crocker & Algina, 1986; Hambleton & Rogers, 1990; Hambleton, Swaminathan, & Rogers, 1991). Classical (Psychometric) Test Theory (CTT) aims at studying the reliability of a test score variable (measurement, test) that maps a crucial aspect of qualitative or quantitative observations into the set of real numbers (Ainsworth, 2005). The test theory is essentially the collection of statistical concepts that formalize and clarify certain questions about constructing and using tests and then provide methods for answering them (Mc'Donald, 1999).

According to (Junker, 2012), CTT provides four methods of evaluating the reliability of an instrument; Split-Half Reliability, Test-Retest Reliability: Parallel Forms Reliability, and Inter-Rater Reliability.

- **Split-Half Reliability** is used to find test score error due to poor test development and calculated the reliability index by applying coefficient alpha, Kuder-Richardson formula 20 (KR-20) or the Spearman-Brown formula.
- **Test-Retest Reliability** is applied to determine test score error due to poor test administration; administer the same test to the same participants on two different occasions and correlate the test scores of the two administrations of the same test.
- **Parallel Forms Reliability** is used to compare the two different versions of the test and calculated through correlation of the test scores of the two versions.
- **Inter-Rater Reliability** is identified the consistency of two separate scorers of the same test, and correlated the ratings from two separate scorers.

Classical test theory is mainly focused on test based information and item statistics. The term “Item statistics” is used to identify item difficulty and item discrimination. CTT collectively considers a pool of examinees and empirically examines their success rate on an item (assuming it is dichotomously scored) (Fan, 1998). This success rate of a particular pool of examinees on an item, well
known as the p value of the item, is used as the index for the item difficulty (actually, it is an inverse indicator of item difficulty, with higher value indicating an easier item). The ability of an item to discriminate between higher ability examinees and lower ability examinees is known as item discrimination, which is often expressed statistically as the Pearson product-moment correlation coefficient between the scores on the item (e.g., 0 and 1 on an item scored right-wrong) and the scores on the total test. When an item is dichotomously scored, this estimate is often computed as a point-biserial correlation coefficient.

The major limitation of CTT can be summarized as circular dependency: a) The person statistic (i.e., observed score) is (item) sample dependent, and (b) the item statistics (i.e., item difficulty and item discrimination) are (examinee) sample dependent. This circular dependency poses some theoretical difficulties in CTT’s application in some measurement situations (e.g., test equating, computerized adaptive testing).

Activity 8.4
Read the different articles based on the “Classical test theory” through use of internet and discuss it with your any classfellow if available or tutors of this course.

8.3.4 Item Response Theory (IRT)

Item Response Theory is a psychometric theory and relates with measurement of latent traits which cannot be measured or observed directly or openly. It is based on the assumptions of probabilistic distribution of examinees’ success at the item level. As its name indicates, IRT primarily focuses on the item-level information in contrast to the CTT’s primary focus on test-level information. The IRT framework encompasses a group of models and the applicability of each model in a particular situation depends on the nature of the test items and the theoretical assumptions about the test items. For test items that are dichotomously scored, there are three IRT models, known as three-, two-, and one-parameter IRT models, and used to analyze the data in psychological assessments and educational testing. (Fan, 1998)

Item response theory is a probabilistic model that attempts to explain the response of a person to an item (Lord, 1980; Hambleton, Swaminathan, & Rogers, 1991). In other words, if a person has a high ability in a particular field, he or she will probably get an easy item correct. Conversely, if a person has a low ability and the item is difficult, he or she will probably get the item wrong. When we analyze item responses, we are trying to answer the question, “What is the probability of a
person with a given ability responding correctly to an item with a given difficulty?" (Culligan, 2015). If a person's ability is greater that the item's difficulty, the probability is more than 50% that the person will answer the item correctly, if the person's ability is less than the difficulty of the item, the probability is less than 50% that the person will answer the question correctly and if the person's ability is equal to the difficulty of the item, the probability is 50% that the person will answer the question correctly. (Marry, 2015)

The important factor that plays important role in popularity of IRT in assessment is to estimate the individual item difficulties and test-taker abilities separately. Another important characteristic of IRT models is local independence: for a given location of test takers on the scale, the probability of success on any item is independent of that of every other item on that scale; this characteristic is the basis of the likelihood function used to estimate test takers’ locations on the scale (Carlson & Davier, 2013). IRT is a model-based measurement model in which latent trait estimates depend on both persons’ responses and the properties of the items and the main purpose of IRT is to create a scale for the interpretation of assessments with useful properties “Scaling” refers to the process by which we choose a set of rules for measuring a phenomenon Creating a “metric” or “scale” for a variable is to systematically assign values to different levels, IRT proceeds in much the same way A meaningful scale is chosen in order to measure subject “ability” or “trait level” The scale can then be interpreted with reference to the characteristics of test items.

Item response theory (IRT) is a theoretical framework organized around the concept of the latent trait. IRT encompasses a set of models and associated statistical procedures that relate observed responses on an instrument to a person’s level of the latent trait. IRT models are used extensively in the study of cognitive and personality traits, health outcomes and in the development of item banks and computerized adaptive testing. A long list of researchers has been working on development and promotion of IRT, some names are given here; (Rasch, 1960; Birnbaum, 1968; wright & Stone, 1979; Lord, 1980; Hambelton, Swaminathen, & Rogers, 1991; King & Bond, 1996; Wu, Hays, Kelly, Malitz, & &Bozzette, 1997; Boardley, Fox, & Robinson, 1999; McDonald, 1999; Embretson & Reise, 2000)

Activity 8.5
Search the ideas of different scholars on “Item Response Theory” using internet and enhance your learning about this theory.
At the end of this unit it can be concluded that test is a tool to assess the students’ learning abilities and skills regardless it is used in formative assessment or summative assessment. Its main purpose is to identify the students’ cognitive developmental process and learning output. A test can be used to ensure the achievement of learning objectives during or post instructional process. The all stakeholders of educational system should do joint efforts to help students in achieving high level thinking and understanding during learning process. In order to maintain the currency and relevancy of the examination, item development is an ongoing process (Marry, 2015). The aim of the reviewing the items is to differentiate the items which are not measuring the learning objectives according to expectations of test. After items are presented on a test, they are analyzed statistically to identify difficulty level of items and discriminating power of items. Standardization attempts to control external factors to the greatest degree possible so that the assessment is a valid measurement tool that produces meaningful results (Zucker, Galindo, Grainger, & Severance, 2004). Students can be compared across the world on the basis of standardized testing. Appropriate standardized procedures improve measurement by increasing consistency and test security. The main purpose of test construction is to develop test items of desired quality, no matter what type of test is used. Classical test theory is mainly focused on test based information and item statistics. The term “Item Statistics” is used to identify item difficulty and item discrimination. According to (Junker, 2012), CTT provides four methods of evaluating the reliability of an instrument; Split-Half Reliability, Test-Retest Reliability: Parallel Forms Reliability and Inter-Rater Reliability. Item response theory (IRT) is a theoretical framework organized around the concept of the latent trait. IRT encompasses a set of models and associated statistical procedures that relate observed responses on an instrument to a person’s level of the latent trait.

8.4 SELF ASSESSMENT QUESTIONS

1. Develop and assemble the classroom test from any textbook of Grade 10th by using seven different formats and arrangements.
2. Develop guidelines to administer any test of Grade 9 and compare it with any manual of standardized testing.
3. Describe briefly the responsibilities of invigilators in standardized testing.
4. Compare the scoring of the objective type test and subjective type test.
5. Choose any test from previous papers of metric exams, and then review its test items qualitatively as well as quantitatively.
6. Describe critically “Item Response Theory (IRT)” & “Classical Test Theory (CTT)”. 

179
INTERNET SOURCES

- https://sites.psu.edu/hybridlearning/what-is-hybrid/
- https://www.online-journals.org/index.php/i-jet/article/view/546
- https://www.igi-global.com/dictionary/the-pivotal-role-of-faculty-in-online-student-engagement-and-retention/20914
- https://www.slideshare.net/stefanstenbom/online-blended-and-webfacilitated-learning-at-kth
- https://www.mghihp.edu/faculty-staff-faculty-compass-teaching/web-enhanced-vs-blended-vs-fully-online-course
- https://eric.ed.gov/?id=EJ1057333
DIFFERENT MODELS OF EVALUATION AND THEIR IMPLICATIONS
CONTENTS

Introduction......................................................................................................................183
Objectives ........................................................................................................................183
9.1 The Four Level model...............................................................................................184
  9.1.1 Responsive Evaluation .......................................................................................184
  9.1.2 Goal-Free Evaluation .........................................................................................184
  9.1.3 Adversary/Judicial Approaches .........................................................................184
  9.1.4 Consumer-Oriented Approaches .....................................................................184
9.2 Utilization-Focused Evaluation ..............................................................................185
9.3 Elements of UFE .....................................................................................................186
9.4 Steps of UFE Framework .......................................................................................187
9.5 Expertise/Accreditation .........................................................................................187
9.6 Participatory/Collaborative ......................................................................................189
9.7 Empowerment .........................................................................................................189
9.8 Organizational Learning ..........................................................................................189
9.9 Theory-Driven ..........................................................................................................189
9.10 Success Case Method .............................................................................................190
9.11 Goal Free Evaluation .............................................................................................190
9.12 Self Assessment Questions .....................................................................................190
  Bibliography ..............................................................................................................192
INTRODUCTION

Evaluation is the process of determining the merit, worth and value of things and evaluations are the products of that process. Evaluation is not the mere accumulation and summarizing of data that are clearly relevant for decision making gathering and analyzing the data that are needed for decision making comprise only one of the two key components in evaluation, a second element is required to get to conclusions about merit or net benefits: evaluative premises or standards. Evaluation has two arms: one engaged in data gathering, the other collects, clarifies and verifies relevant values and standards (Smith, 2001).

OBJECTIVES

After studying this unit, the prospective teachers will be able to:

* define what evaluation is.
* understand the evaluation as a whole process.
* delineate various stages of evaluation.
* to implement any type of evaluation.
* determining when and where to use various evaluation models and approaches
* report the results of evaluation
9.1 THE FOUR-LEVEL MODEL

This approach is most often used to evaluate training and development programs (Kirkpatrick, 1994). It focuses on four levels of training outcomes: reactions, learning, behavior, and results. The major question guiding this kind of evaluation is, “What impact did the training have on participants in terms of their reactions, learning, behavior, and organizational results? Evaluation approaches are conceptually distinct ways of thinking about, designing and conducting evaluation efforts.

9.1.1 Responsive Evaluation

This approach calls for evaluators to be responsive to the information needs of various audiences or stakeholders. The major question guiding this kind of evaluation is, “What does the program look like to different people?”

9.1.2 Goal-Free Evaluation

This approach focuses on the actual outcomes rather than the intended outcomes of a program. Thus, the evaluator has minimal contact with the program managers and staff and is unaware of the program’s stated goals and objectives. The major question addressed in this kind of evaluation is, “What are all the effects of the program, including any side effects?”

9.1.3 Adversary/Judicial Approaches

These approaches adapt the legal paradigm to program evaluation. Thus, two teams of evaluators representing two views of the program’s effects argue their cases based on the evidence (data) collected. Then, a judge or a panel of judges decides which side has made a better case and makes a ruling. The question this type of evaluation addresses is, “What are the arguments for and against the program?”

9.1.4 Consumer-Oriented Approaches

Consumer-Oriented Evaluation Developed by Scriven. Much like needs assessments and the expertise-oriented model of evaluation. It helps in inform decisions on what to purchase or trade. This approach judge the quality of something, establish value, merit or worth of a product. Audience is broader. The purchasing public and they are not known to the evaluator.
The emphasis of this approach is to help consumers choose among competing programs or products. Consumer reports provide an example of this type of evaluation. The major question addressed by this evaluation is, “Would an educated consumer choose this program or product?”

This approach is important to:
- Identify the criteria correctly
- Develop standards to judge those criteria
- Collect data
- Synthesize information to make a final judgment

**Consumer-Oriented Examples:** Consumer Reports, US News and World Report

### 9.2 UTILIZATION-FOCUSED EVALUATION

According to Patton (1997), “utilization-focused program evaluation is evaluation done for and with specific, intended primary users for specific, intended uses”. As such, it assumes that stakeholders will have a high degree of involvement in many, if not all, phases of the evaluation. The major question being addressed is, “What are the information needs of stakeholders, and how will they use the findings?”

Utilization-Focused Evaluation (UFE), developed by Michael Quinn Patton, is an approach based on the principle that an evaluation should be judged on its usefulness to its intended users. It is a decision-making framework for enhancing the utility and actual use of evaluations. Utilization-focused evaluation is concerned with how real people in the real world apply evaluation findings and experiences the evaluation process. Therefore, the focus in utilization-focused evaluation is on intended use by intended users. In any evaluation there are many potential stakeholders and an array of possible uses. Utilization-focused evaluation requires moving from the general to abstract. No evaluation can be value-free, utilization-focused evaluation answers the question of whose values will frame the evaluation by working with clearly identified, primary intended users who have responsibility to apply evaluation findings and implement recommendations (Newman, 1995). In essence, utilization-focused evaluation is premised on the understanding that evaluation use is too important to be merely hoped for or assumed. Use must be planned for and facilitated.

Utilization-focused evaluation is highly personal and situational. The evaluation facilitator develops a working relationship with intended users to help them determine what kind of evaluation they need. This requires negotiation in which
the evaluator offers a menu of possibilities. Utilization-focused evaluation does not depend on or advocate any particular evaluation content, model, method, theory, or even use. Rather, it is a process for helping primary intended users select the most appropriate content, model, methods, theory, and uses for their particular situation (Patton, 2008).

In considering the rich and varied menu of evaluation, utilization-focused evaluation can include any evaluative purpose (formative, summative, developmental), any kind of data (quantitative, qualitative, mixed), any kind of design (e.g., naturalistic, experimental) and any kind of focus (processes, outcomes, impacts, costs and cost benefit, among many possibilities). Utilization-focused evaluation is a process for making decisions about these issues in collaboration with an identified group of primary users focusing on their intended uses of evaluation (Wye, 1995).

While concern about utility drives a utilization-focused evaluation, the evaluator must also attend to the evaluation's accuracy, feasibility and propriety (Joint Committee, 1994). End to the evaluation's accuracy, feasibility and propriety (Joint Committee, 1994). Moreover, as a professional, the evaluator has a responsibility to act in accordance with the profession's adopted principles of conducting systematic, data-based inquiries; performing competently; ensuring the honesty and integrity of the entire evaluation process; respecting the people involved in and affected by the evaluation; and being sensitive to the diversity of interests and values that may be related to the general and public welfare (Shadish, Newman, Scheirer, & Wye 1995).

9.3 ELEMENTS OF UFE

UFE has two essential elements.

- The primary intended users of the evaluation must be clearly identified and personally engaged at the beginning of the evaluation process to ensure that their primary intended uses can be identified.

- Evaluators must ensure that these intended uses of the evaluation by the primary intended users guide all other decisions that are made about the evaluation process.

UFE can be used for different types of evaluation (formative, summative, process, impact) and it can use different research designs and types of data.
9.4 STEPS OF UEF FRAMEWORK

There are 17 Step UFE Frameworks

- Assess and build program and organizational readiness for utilization-focused evaluation
- Assess and enhance evaluator readiness and competence to undertake a utilization-focused evaluation
- Identify, organize, and engage primary intended users: the personal factor
- Situation analysis conducted jointly with primary intended users
- Identify and prioritize primary intended uses by determining priority purposes
- Consider and build in process uses if and as appropriate
- Focus priority evaluation questions
- Check that fundamental areas for evaluation inquiry are being adequately addressed: implementation, outcomes, and attribution questions
- Determine what intervention model or theory of change is being evaluated
- Negotiate appropriate methods to generate credible findings that support intended use by intended users
- Make sure intended users understand potential methods controversies and their implications
- Simulate use of findings: evaluation's equivalent of a dress rehearsal
- Gather data with ongoing attention to use
- Organize and present the data for interpretation and use by primary intended users: analysis, interpretation, judgment, and recommendations
- Prepare an evaluation report to facilitate use and disseminate significant findings to expand influence
- Follow up with primary intended users to facilitate and enhance use
- Meta-evaluation of use: be accountable, learn, and improve.

9.5 EXPERTISE/ACCREDITATION APPROACHES

The accreditation model relies on expert opinion to determine the quality of programs. The purpose is to provide professional judgments of quality. The question addressed in this kind of evaluation is, “How would professionals rate this program?”

It is the oldest type of formal evaluation. It relies on professional expertise to judge the quality of following things;
- An institution,
- Program,
Product,
✓ Activity

Types of Expertise-Oriented Evaluations

Following are the types of expertise oriented evaluation.
- Formal review system
- Informal review system
- Ad hoc panel review
- Ad hoc individual review

9.5.1 Formal Review System

This system mostly used in accreditation. It examine existing structure of organization and also examine published standards. This review system follows a specified schedule. It also plays an important role to increase emphasis on outcomes. It uses the opinions of multiple experts status of those being evaluated is affected by the results.

Formal Professional Review System includes;
- The structure or organization established to conduct periodic reviews of educational endeavors,
- Published standards,
- Pre-specified schedule,
- Opinions of several experts
- Impact on status of that which is reviewed

9.5.2 Informal Review

The informal system of review used primarily for evaluations that lack published standards or follow a specified review schedule. It use the multiple reviewers status of those being reviewed is affected by results.

Examples: Peer review of articles, theses or dissertation committees.

9.5.3 Ad Hoc Panel Reviews

Ad hoc panel reviews occur at irregular intervals when circumstances demand. These reviews not related to institutionalized evaluation or standards. Usually one-shot evaluations prompted by a particular, time-bound need for evaluative information.

Examples of ad hoc panels: Panels to develop standards, Funding agency review panels, Blue ribbon panels.
9.5.4 Ad Hoc Individual Reviews

Ad hoc individual reviews, Review of any entity by an individual selected for his/her expertise. This review use usually to judge value or make recommendations

**Example:** Employment of a consultant to review an educational, social, or commercial program

9.6 PARTICIPATORY/COLLABORATIVE EVALUATION

The emphasis of participatory/collaborative forms of evaluation is engaging stakeholders in the evaluation process, so they may better understand evaluation and the program being evaluated and ultimately use the evaluation findings for decision-making purposes. As with utilization focused evaluation, the major focusing question is, “What are the information needs of those closest to the program?”

9.7 EMPOWERMENT EVALUATION

This approach, as defined by Fetterman (2001), is the “use of evaluation concepts, techniques, and findings to foster improvement and self-determination” (p. 3). The major question characterizing this approach is, “What are the information needs to foster improvement and self determination?”

9.8 ORGANIZATIONAL LEARNING

Some evaluators envision evaluation as a catalyst for learning in the workplace (Preskill & Torres, 1999). Thus, evaluation can be viewed as a social activity in which evaluation issues are constructed by and acted on by organization members. The major question in this case is, “What are the information and learning needs of individuals, teams, and the organization in general?”

9.9 THEORY-DRIVEN EVALUATION

This approach to evaluation focuses on theoretical rather than methodological issues. The basic idea is to use the “program’s rationale or theory as the basis of an evaluation to understand the program’s development and impact” (Smith, 1994, p. 83). The major focusing questions here are, “How is the program supposed to work? What are the assumptions underlying the program’s development and implementation?”
Steps:
Following are the steps of theory driven evaluation.
- Engage relevant stakeholders
- Develop a first draft of program theory
- Present a draft to stakeholders for discussion, reaction, and input
- Conduct a plausibility check
- Communicate findings to key stakeholders
- Probe arrows for model specificity
- Finalize program impact theory

9.10 SUCCESS CASE METHOD

This approach to evaluation focuses on the practicalities of defining successful outcomes and success cases (Brinkerhoff, 2003) and uses some of the processes from theory-driven evaluation to determine the linkages, which may take the form of a logic model, an impact model, or a results map. Evaluators using this approach gather stories within the organization to determine what is happening and what is being achieved. The major question this approach asks is, “What is really happening?”

9.11 Goal-Free Model of Evaluation (Scriven)

The evaluator purposely avoids becoming aware of the program goals. Predetermined goals are not permitted to narrow the focus of the evaluation study. Goal-free evaluation focuses on the actual outcomes rather than intended program outcomes. The goal-free evaluator has minimal contact with the program manager and staff. This type of evaluation increases the likelihood that unanticipated side effects will be noted.

9.12 SELF ASSESSMENT QUESTIONS

Subjective types:
1. Define the term evaluation and elaborate the importance and scope of evaluation.
2. State different models of evaluation and explain each type with example.
3. Apply Kirkpatrick’s four-level approach to an evaluation case scenario.
4. Consider and discuss the advantages and disadvantages of using these four-level approaches to evaluating training and development programs.
5. Discuss the reasons why the Kirkpatrick approach has been so popular in the training and development field.
6. Present the reasons for selecting a particular evaluation approach.
7. How does one choose an evaluation approach? What criteria drove your choice?

**Objective types:**

Make the following statements true or false:
- Evaluation is about gathering data.
- Training data is used in model evaluation.
- Evaluation is only about measuring a phenomenon.
- It assesses quality through quantitative or qualitative means.
- Evaluation is a goal oriented activity.
- The Four-Level Model is most often used to evaluate training and development programs.
- The accreditation model relies on expert opinion to determine the quality of programs.
- Utilization-focused evaluation is highly personal and situational.
IBLIOGRAPHY


