UNIT

19 Online/Web-Based Student Assessment

Introduction

If we talk about the testing in a conventional setup, student evaluation means that the student has to write an examination of two-three hours, which will test his level at a given point of time and in a locally similar type of environmental setting. The word locally is used here because, unlike Distance Education,Conventional Education is imparted in one campus only. These tests measure the knowledge and skills considered important in the learning. However, the goal of these evaluations is to just ascertain that at the time of taking the examination the student should be able to exhibit his performance on the concepts asked in the question paper. There is an immense importance of achieving intended learning outcomes. However, these kinds of evaluations are restricted to student achieving marks or grades that are as per the pre-defined pass level. These evaluations are, in their entirety, summative in nature. But what is happening today is, rather than using testing as a learning tool, students also become interested in getting taught for the test material itself. To be significant, the act of evaluation must enhance the learning process and thus comes the concept of continuous evaluation component, which is formative evaluation, intended to provide feedback to students on the errors, they have committed in understanding the concepts. In a conventional setup, it is easier to monitor the student's performance and provide feedback. However, in Distance Education setup, it is extremely difficult to do so. With the emergence of institutions like IGNOU, which has an enrollment of more than 10 lakhs of students, it is increasing becoming imminent that technology comes in a big way to the rescue. This is the main reason of our interest in online/web-based assessment which on the one hand, perform both summative and formative functions of evaluation and on the other hand remove the biases which are prevalent in human evaluation.

This unit is intended for the teachers who are using or are planning to use computers in their classrooms for enhancing the quality of the learning of their learners and also finding it difficult to manage the load of growing numbers. Computers have received their due importance in instructional activities, but there is a lot of scope for developing its role in assessment and feedback.

Learning Outcomes

After reading this unit, you would be able to:

- define the concept of online and web-based evaluation
- identify the issues involved in generating an electronic test
- discuss the theories on which the essay-evaluators work, and
- discuss other forms of evaluation which can be used in web-based evaluation.

Computers in student evaluation

Many of the recent advances made in educational measurement and evaluation could be attributed to computers. There is a continuous update in both hardware and software and hence there is an emergence of many newer applications. People have ralized that application of these technologies can enhance the quality of learning experience. However, one must admit that the use of computers in testing is much less than its use in instructional activities. The use of computers in evaluation has probably risen due to the enhanced time, money and effort spent. Technology can be used for assessment puposes at multistages, from the management of the assessment information to a fully automated assessment system. In recent times, education world has seen emergence of several areas in computer-based assessment. There has been a growing interest and increasing practical experience in the use of computer to deliver objective tests. Objective testing is often taken to imply the use of multiple-choice questions. Many packages are available which are designed for the electronic delivery of objective tests, all of which support the delivery of a variety of question types.

In addition to using electronic packages to create unique tests, it is possible to use the computer to generate different tests automatically. Computerized delivery from a bank of questions or of randomly generated problems, supported by automated marking and feedback to the student is a flexible and efficient method of providing formative assessment, particularly when factual knowledge is an importnat component of the course.

In recent times the education world has seen the emergence of two efforts at using computers to evaluate the content of essays.

These are:

- Intellignt Essay Assessor (IEA), developed by Peter Follz of Univerity of Colorado at Boulder
- Graduate Management Admission Test (GMAT), which employs a system called E-Rater

Reasons for their development were the dificulty in managing the load of grading the essays and the tendency of the evaluators to shift the focus of assessment to multiple-choice questions. In several studies IEA agreed with each human grader as often as they agreed with each other. These approaches not only save time but also provide consistency a computer can deliver. They are also helpful in training of evaluators and for giving feedback to the students. We have seen that these models are defined for Languages, Psychology, and Education, etc. But the interesting question is how do we use computerized models for Mathematics, Computer Science and other Sciences and Management subjects? Well, we have to go back to the Basics. If we go into details, there are 51 learning theories, 18 learning domains and 18 learning concepts. These theories provide an interesting opportunity for the teachers to assess their student's learning in a purposeful manner.

There are immense benefits of using online and web-based evaluation. The first and foremost benefit is in the increased frequency of evaluation that performs both formative and summative functions. Putting quizzes and assignments on web and enabling the student to do a sort of self-assessment can perform the formative function. The teacher can monitor this self-assessment and feedback can be offered to a single or a group of students as the requirement is. It can also aid in summative function as it can considerably reduce time spent in marking, which in turn can be utilized for other meaningful activities. The teachers can declare the results with a click of a button, which will eventually aid the students in their planning of further studies.

Electronic delivery of objective tests

The prevalent evaluation system, in most of the institutions, has two components i.e. the continuous assessment component carrying a weight of 30% and a term-end component carrying a weight of 70%. The former component is further broken into two components based on the nature of the assignments i.e. Computer-marked assignment (CMA) and Tutor-marked assignment (TMA). Computer marked assignments are generally multiple-choice questions (MCQ). However, objective type tests can have a lot of other formats like fill in the blanks, word entry or number entry, identifying a segment etc.

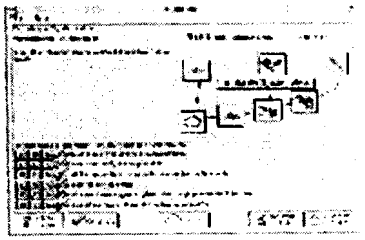
Several packs are developed and are available for electronic delivery of objective tests. These packages offer a huge variety of questions formats. These packages are capable of giving feedback and doing assessment also. You will, in this section read about a few of them. To begin this discussion it is pertinent to mention CAPES which was developed by the National Informatics Centre, India in 1993. CAPES is called Computer Aided Paperless Examination System. This system was based on optical memory technology, especially CD-ROM where large amount of data pertaining to optimally designed questions banks and answers can be stored. The system then with the use of computer networks allows conduct of examination with minimum requirement of manpower for supervision and maintaining security of the examination. The system was also capable of delivering similar tests for all examinees in terms of difficulty levels, course coverage, and biases etc. by the use of stratified sampling techniques. A test of CAPES on 6800 examinees distributed over 20 days and 18 cities demonstrated that the system has not introduced any bias in the randomization process and the time spread also does not have any biases over the scoring ability of examinees taking the test towards the end duration as compared to those at the beginning.

IGNOU has also developed an online and on-demand model called IGOLEX, which is in the testing stage at the moment. IGOLEX is defined as the IGNOU Online Examination system. In this system, it is planned that every student who wishes to appear for an examination has to register for an examination. On the basis of his registration the student would be given a password for logging on to the site. The examination would produce a different paper of equivalent standard each time an examinee logs on. The question bank in this system is prepared with different characteristics like item difficulty, expected time taken, course unit, block etc. embedded in them. As soon as an examinee submits its work the system evaluates the work. The marks are reported to the student and simultaneously get recorded in the server at IGNOU for entry into the respective student grade card.

There are several other packages that are exclusively meant for delivering objective tests. We would discuss two of them, i.e. *Examine* and *Question Mark. Examine* is a multiple-choice delivery system for Windows developed by University of Nottingham. *Examine* can be used not only as an assessment tool but as a teaching tool also. It can help the learner to do a self-assessment and revision. The questions can be put into a text file that can be printed if one wants so. The *Examine* system consists of two components: *Examine* is delivery engine that is proposed for the student use and *Quiz Maker* is a tutors tool that enable teacher to generate quizzes from the available questions. *Examine's* question bank is segregated into many question folders as per the chapters in the book. These questions are of many types like Multiple-choice questions, Multiple selections, True/False, Numeric and Comments. It offers the flexibility to offer random quizzes by selecting a random number of questions from each chapter.

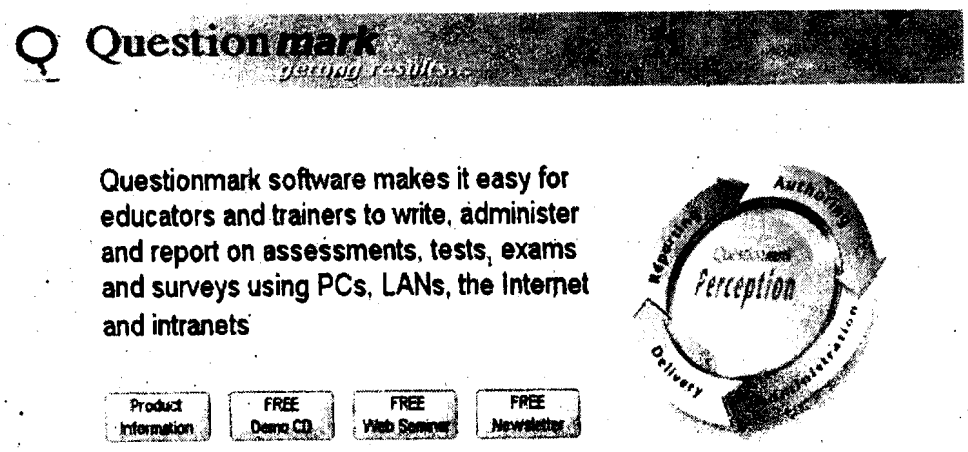
Examine

Examine **is a multiple-choice authoring and delivery system for use either as an adjunct to courseware or as a standalone means of on-line self-assessment. A number of question types are supported, as well as graphics and multimedia. There is also support for the Microcosm hypermedia system developed by Southampton University. Recently *Examine* has been licensed to McGraw Hill for distribution with a large collection of economics questions under the name of *Begg Examine* (this will form part of a teaching pack that supports the book Economics by David Begg et al).**



Sourece: http://ibis.nott.ac.uk/software/examine.html

Questionmark Perception is another software that helps the teachers to write, administer and report on assessments, quizzes and surveys using computers. *Perception for Web* is a package for conducting assessments using the Internet and Entrants. *Questonmark* is an authoring system with many question types. It has a good feature that enable one to deliver the assessment in an Off-line mode but still capable of storing the answer. It has integrated Sound and Video with the questions and their answers so that it enhances the effect of assessment and overall learning. These clips can be configured to auto-run. Each question contains a text note and metadata. Apart from the question type discussed above it also contains question types like matching, pull-down list, rank in order and drag & drop kinds.



Sourece: http://www.questionmark.com/us/home.htm

So you can see that virtually all the tools are readily available for you to deliver online electronic tests. However, this area is yet to develop itself for the numeric and other disciplines that use special notations for the answer. However this can be overcome partially by the use of Drag & Drop feature. Thus online evaluation helps you by offering interesting possibilities that are not available with conventional paper-based systems. First, by creating a computerized question bank, you can generate as many different

questions as you want at the same time and with equivalent standard. It can also save your precious time that is otherwise used in making the question paper again and again. Second, by making an instant evaluation through computer you can give immediate feedback by storing the answer also in the question bank. Third, Online/Web-based evaluation through its ever readiness can enable you to offer on-demand examination to the students, i.e. giving them the flexibility to take the exam as and when they feel their learning is complete. Lastly, you can record the results immediately, as it is easier to analyze the group responses and do test and item analysis to provide feedback for future examinations.

Self-assessment

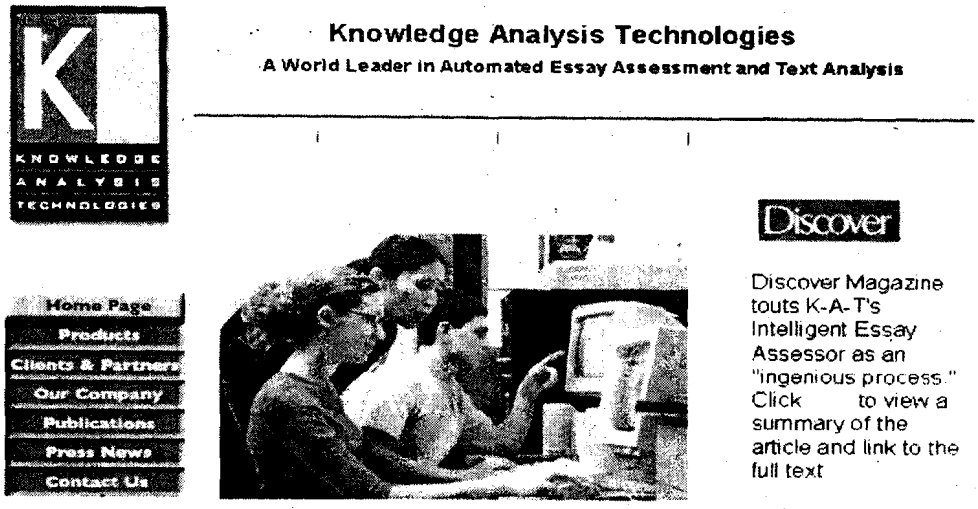
1. *List the disadvantages of using objective tests only in the online assessment.*
2. *Mention any other evaluation tools that can be used in online assessment.*

Possibilities in subjective tests

Technology has made it possible to think beyond objective test. Restrictions in including only right/wrong kind questions in online evaluation systems are no more in place. There do exist some packages that pop up a hint when a student is unable to answer a question. Some packages also offer a feedback on what kind of error is committed and what should be the correct answer. However, these packages will take answer input as a word or a numeral only. Can you think of a situation in which the computer grades an essay as you do? As we have mentioned it earlier, a beginning has been made in this direction and time is not very far when you will be equipped with customized software to evaluate students in your very specific course through the computer. The constraints are none but one: The student has to key his own answer in the computer. In 1998, a Professor of Psychology at University of Colorado at Boulder completed work on the software after a hard effort of 10 years. The software shifted its focus on the content from writing and found out that these two are highly correlated. As a teacher you have to feed information into the computer from the textbooks and other sources. The computer in turn forms relationships between the words. As a teacher you are again expected to give computer a set of model answers that are graded by the human evaluators, which would eventually provide the computer a basis for comparison. It is not as simple as it looks but then once you are through you would be able to evaluate thousands of essays with the click of a button. The software also provides feedback on what the student left out in the answer besides giving a grade to the answer. We will discuss two of such approaches in detail to give you a feel of how these softwares work.

Intelligent Essay Assessor (IEA) is based on a statistical approach for analyzing the essays and content information. This statistical approach is called Latent Semantic Analysis, which is used primarily to extract semantic content of the text. Many scientists have proved that latent semantic analysis captures the similarity of meanings expressed. This technique is successful in giving judgements very close to the human judgements. According to the developers of the IEA, the semantic analysis focuses on the conceptual content of the essay, i.e. on the correctness and completeness of the content, the soundness of the arguments and the fluency and comprehensibility of the writing. This was a deviation from the earlier approaches, which were more mechanical in their approach of giving importance to grammar, punctuation and spelling.

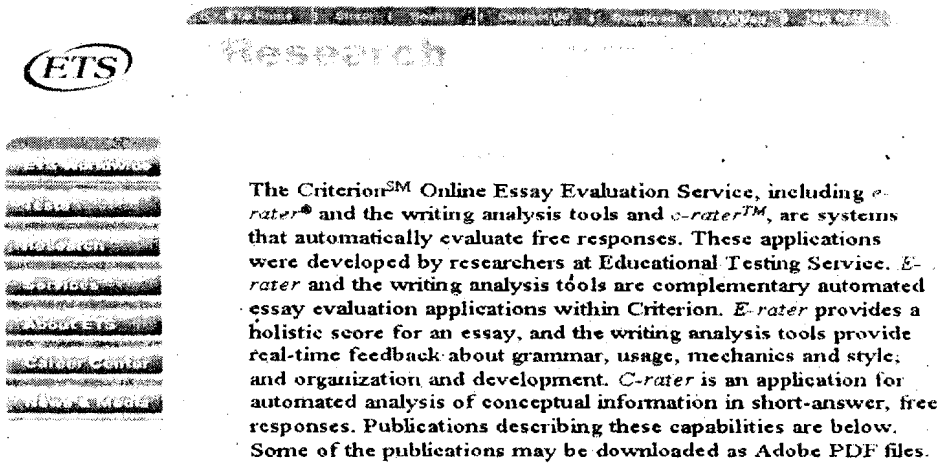
You will get to know about Latent Semantic Analysis and other techniques later. This intelligent Essay Assessor has been used in a course of Psycholinguistics at New Mexico State University as a web based application. One can visit the website http://psch.nmsu.edu/essay for a demonstration of this package. This package basically permits the students of that course to log on the web and submit their essay to get a feedback on the missing points in their essays. After going through the feedback, students



Sourece: http://www.knowledge-technologies.com/

can revise their essays and resubmit the work. The main benefit of this process is that on one hand the software is performing the function of summative evaluation and on the other hand it is also doing the function of formative evaluation.

E-rater is another software worth mentioning in the computerized essay evaluators segment. The Educational Testing Service uses it for the Graduate Management Admission Test for testing the analytical writing assessments. E-rater is used as a second rater replacing the second human rater and thus reducing large costs on evaluation. This system has been developed after 5 years of rigorous efforts in using the advanced computational linguistics techniques. Apart from GMAT test, e-rater was also applied on Test of English as a Foreign Language (TOEFL) test. E-rater follows the scoring guide used by expert human evaluators for scoring. This scoring guide has a six-point scoring scale. It checks the essay for the argument structure, syntactic structure and vocabulary structure. The software is based on three general classes of features: syntactic, rhetorical and topical content features. The features are extracted from the essay texts and qualified using computational linguistics techniques.



Source: http://www.ets.org/research/erater.html

Methodologies of essay evaluators

In the earlier sections, you have learnt about the two efforts in evaluating the essays through computer and thus adding a new dimension to web-based student assessment. Since the theories behind these methodologies are a new in the educational and a little statistical in nature, it is worth giving an overview of them. This will enable you to appreciate these efforts better.

a) **Latent Semantic Analysis** It is a statistical model of word usage that permits comparisons of the semantic similarity between pieces of textual information. LSA assumes that there is a latent or concealed structure in the pattern of word usage in the essays and statistical can be used to eliminate this masked thing. The method generates a depiction of words that are used in similar contexts and are more semantically associated. LSA generates a matrix of occurrences of each work in each essay and then decomposes these matrices into a set of hundreds of factors. Since the number of factors will be much smaller than the number of unique words, words will not be independent. If two terms are used in same context, they will have similar factors. Matching is then done between two pieces of textual information, even if they have no words in common.

b) **Syntactic Structure Analysis** Syntactic variety is an important feature in evaluating essays. It defines the ratio and quantity of types of sentences, types of clauses and use of verbs. This analysis parses through each sentence in the essay and quantifies these features. This parsing is done on the basis of Microsoft's Natural Language Processing tool.

c) **Rhetorical Structure Analysis** If an examinee is able to demonstrate that the essay he developed, organized the ideas logically and has connected them well, then he ought to get a good score. This analysis quantifies the evidence of organization of the essay. It quantifies the cue words and other structures of the essay. Examples of cue words that are used to define a relation are given by the developers are: In summary, In conclusion, etc.

d) **Topical Content Analysis** A good essay resembles other good essays in its use of vocabulary use of patterns. This analysis evaluates the topical content of an essay by comparing it with the other model essays that are graded by the human evaluators. It uses two different measures of content similarity. First, one is based on the vocabulary use in the essay as a whole and other is based on specific vocabulary content of the argument found in the essay.

Implications for education

The computerized essay evaluators have the following implications for education:

- *Reliability Regulator* Many teachers grade an essay and then use a computerized essay evaluator to ascertain the correlation. Only computer can offer unbiasedness, consistency and robustness.
- *Mass Evaluation tool* Computerized essay evaluators can be used for a large number of examinees with a less cost and time under the identical evaluation yardsticks.
- *Friendly On-Demand Tutor* Computerized essay evaluators can work as a friend philosopher and guide to the students by offering comments and their work and thus improving the quality of their writing. This can be of great help in distance education where the student is fairly distant from the teacher.
- *Pre-Testing Tool* Computerized essay evaluators can work as a great tool for the students to know beforehand about their preparations for the final term-end exam. Students can write and review their essays at the end of each chapter and then assess the areas on which they have to work more.

- *Second Rater* Computerized essay evaluators can work as a second rater to cross check the major differences, if any, with the human evaluators. This is especially useful in the high stake assessments such as in GMAT and TOEFL.

In short, computerized essay evaluator presents a good tool for assessing the quality of the essay in question. These can be applied in the classrooms and in distance learning for deriving respective benefits. The need at the moment is to generate the artificial intelligence in varied subjects.

Self-assessment

3. *Mention an area in which computerized essay evaluators cannot perform well. Give reasons for your choice.*
4. *Take a test on some students in your class. Get it evaluated by 4-5 evaluators. Try to find the correlation between them. What did you observe?*

Other tests suitable for online/web-based assessment

Having dealt with objective and subjective tests, you might be wondering as to what other tools are left to complete the discussion on web-based student assessment. Well, there are many which you might be using in your classroom environment. Technology has made it possible to adapt them to the web-based evaluation model. You will learn about direct presentation tools, electronic seminars and conferencing, authoring tasks and simulations in this session.

Distance education can be imparted wonderfully by making it collaborative and networked. It is here that the *Direct Presentation tools* come into picture. Along with the growing use of computers in institutions, it is not very difficult for the students to generate their work on a word-processor or as a presentation in Power Point. After completing the work a student can put it on the web or the local area network so that peers see it and add a comment to it. The teacher as a moderator can see all this work and offer comments on it. *Common Space* is a good software for this activity. . It can be used by a group of individuals who offer comments on an individual's work. These comments can be integrated to revise the work and hence improve the quality.

Electronic Seminars and Conferencing is another area where computer and web-based technologies offer a tool for the assessment. You must have seen various discussion groups where people share their view through emails. A teacher can work as the moderator of the discussion group and offer comments to the group members (essentially students of that course). Conferences are also held electronically now a days. In this conference one submits his work to a group of established experts through email for consideration and the review process.

Authoring Tasks Simple authoring tasks can be given to the students, which leads them to the development of small packages. For example, if you are teaching a course on educational multimedia, then the best way to assess the deeper understanding of the student is by asking him a small multimedia package. Similarly, creation of a lecture-note, creation of a web page in HTML, etc. can give a practical touch to your lectures. This is discussed here because computers help in a big way in execution of these works and web helps in submitting them for an assessment.

Advantages of online/web-based student assessment

McCormack and Jones (1998) have highlighted many advantages of the online/web-based student assessment. Assessment created for online/web-based assessment do require a lot of time when established for the first time, but in the long run proves beneficial by

reducing the time, cost and efforts. They enable generation of a true feedback to the student instantaneously, which is otherwise impossible especially in the context of Distance Eduction, where it takes more time to evaluate assignments and providing feedback on them. A study of the turn-around time for the assignments showed that sometime it takes months for the assignments to reach back to the student. Web-based assessment eliminates the with-in and between biases of the human evalutors and thus provides consistency and reliability which is impossible for a human evaluator. It can help in automatically recording the results as and when an assessment is completed. Remember, this generates an error-free data, which is otherwise scarce, when you deal with several thousands of students at a time. If the assignments and their assessment are done through web, a lot of efforts can be saved in dispatching and collecting them through conventional methods like postal service. It also reduces the chances of assignments missing and delay in results.

Offline use of computers in student assessment

When we talk about Online/Web-based student assessment, we are not necessarily talking about the delivery through web or a local area network (LAN). You will be surprised that most of these software work in off-line mode also. The only difference is that delivery of such a test in offline mode requires a little bit of adjustment, e.g. in place of web or LAN, one can use postal or personal delivery of the CD-ROM containing the computerized question banks from which any number of question papers can be made through random number generation. IGOLEX has such a feature that when a student would enter his enrolment number and password based on some formula, the computerized question bank will generate a unique question paper for him. The use of question bank feature can also be seen with QUIZMAKER tool of EXAMINE. Also you would recall that QUESTIONMARK has a feature that enables one to deliver the assessment in an off-line mode that is still capable of storing the answer.

Apart from generating the test, computers can do many other things in the off-line mode, e.g. analysis and recording of the awards received by the students, recording the most prevalent errors committed by the students and the type of feedback given when a particular error is committed. These are some of the tasks a computer can do while in off-line mode. E-rater and IEA can also work in off-line mode. If this is so, a question that arises is, "what is the difference between off-line and on-line mode then?" One can answer it by saying that on-line mode extends the range of possible receptors of the assessment, say, if in on-line mode thousands of students can be assessed, in off-line mode you have to restrict it to a classroom or a college. The second difference is that in on-line mode you can load the system on one computer and run the show on many networked computers at the same point of time, but in off-line mode you have to load the system on computer on which you would like to carry out an assessment. Another difference is on the count of recording the work of students. In online mode, system generates automatic delivery of the responses of examinees but in off-line mode, one has to record them separately and then deliver it by alternative means.

There are many other differences, can you think and try to list a few of them?

Summary

We are still in the infancy stage when we talk of online/web-based student assessment. The progress in imparting on-line learning is phenomenal, but on the online assessment front a lot is still to be done. But the good thing is that tools are available and are in abundance. The need is the change in the attitude. Although it is said again and again that a university's reputation is built by the element of sanctity in its evaluation procedures, there has been a slow progress in adopting web-based assessment techniques, which can assure this.

In this Unit, you have broken the myth that online/web-based evaluation is nothing but objective-type multiple-choice questions. You have seen that virtually all the questions type which exists in conventional testing are capable of being handled through web. These systems on one hand enhances the learning opportunities for the students and on the other hand provides tools for the teacher to do much more rather than wasting time on redundant assessments.

Suggested readings

Burstein Jill, Karen Kukich, Lisa Braden-Harder, Martin Chodorow, Shuyi Hua, Bruce Kaplan, Chi Lu, James Nolan, Don rock and Susane Wolf (1998a), Computer Analysis of Essay for automated Score Prediction: A Prototype automated scoring system for GMAT Analytical Score Prediction. (RR-98-15). Princeton, NJ: Educational Testing Service.

N. Seshagiri (1993), Capes-Computer aided paperless examination system Part-I: Theoretical Basis, INFOTEX'93, Bangalore, India.

M.V.V.S. Reddy, N. Seshagiri (1993), Capes-Computer aided paperless examination system Prt-II: Field test results, INFOTEX'93, Bangalore India.

Foltz, P.W. (1996), Latent Semantic Analysis for the Text-based research. *Behaviour Research Methods, Instruments and Computers. 28(2), 197-202.*

McCormack and Jones (1998), *Building a web-based eduction system,* Wiley Computer Publishing, New York, pp. 236-7.

Answers to self-assessment

1. Some important limitations are:
   a) They are unable to measure higher-level skills. They only measure factual knowledge.
   b) At the initial stage the implementation could be time consuming and tedious.
   c) Setting-up of the infrastructure for the conduct and Hardware/software reliability is highly essential.
   d) Students who are not familiar with computers will be in a disadvantageous situation.
2. Some other interesting possibilities could be checklists e.g. to undertake a project work, a student ticks all the steps he/she has to go through for completion of the project. Image mapping e.g. to identity hardware components of a computer the examinee select an image which resembles with the component typically.
3. Computerized essay evaluators are constrained to work with subjects, which are mathematical in nature. It is also not suitable for evaluating numerical. The subjects in whom an answer can be answered in many different ways can also make it difficult to generate intelligence for the computer.
4. You must have observed that there could be a lot of variability between the scores of different evaluators. In fact if you yourself evaluate these answers at different points of time. you will still find variability.