CHAPTER 2

REQUIREMENTS OF AN EFFECTIVE COMPREHENSIVE COMPUTER-ASSISTED ASSESSMENT TOOL

2.1 INTRODUCTION

This chapter will focus on the technical characteristics and operational and pedagogical requirements of an online assessment tool. The second part of the chapter will provide a critical overview of quantitative measurement values that are required to ensure validity, reliability and fairness of items included in an item bank.

Broader issues relating to assessment will be examined, where after the focus will shift more specifically to Computer-Assisted Assessment (CAA).

2.2 ASSESSMENT

Assessment practices are widely discussed in the literature and research articles (Smith, Scholten, Russell & McCormack 1997:78; Sutherland & Peckham 1998:98). Frequently asked questions are: Why assess? What is the main purpose of assessment? What should be assessed? What should be the scope of assessment? Who is to be assessed, when and by whom? How should we assess?

2.3 WHAT IS ASSESSMENT?

Assessment is the process of collecting information. Salvia and Ysseldyke (2001:3) regard assessment as more than just the collection of information. They regard it as collection with purpose. The single most difficult aspect of the education process is how we assess what we have taught. Educators' assessment of students forms an integral part of the teaching and learning process. Assessment can only take place if tests in different forms are administered to students, therefore its purpose is to make an assessment of the students' attainment of knowledge, their acquisition of skills and competence, and finally their performance.

According to Bresciani (2004), the sole purpose of a test is to make a measurement. Assessment is very much a process of measurement, whether the outcome is used for baseline, diagnostics, formative or summative purposes. When measurement is taken, in whatever form, a score is obtained. The score that is obtained forms the important part of assessment, because this score determines the outcome of the assessment, the decisions that are to be made regarding students' progress, curriculum changes and the evaluation of a course as a whole. Although a score is obtained from a test, its analysis is much neglected.

The main aim of this study was to show that scores obtained from tests administered to students are only trustworthy if they are properly analysed. Before a score can be analysed the purpose of assessment must be considered.

2.4 PURPOSE OF ASSESSMENT

Assessment cannot take place in isolation; it is an integral part of the teaching process. For teaching to take place people are needed with the necessary skills and competencies to teach. In order to become educators, people go through a great deal of effort to obtain the required certification to be able to teach. It is a fact that most educators teach because they like to teach (Popham 1995:1). Although educators like to teach, they rarely like to test (Popham 1995:1).

Can we make the assumptions that good educators are good assessors? It is the view of Popham (1995) that educators who can test well will be better educators. Effective testing will enhance an educator's instructional effectiveness; that is, it will assist the educator in making a decision on any adaptation that needs to be made in the way instruction is taking place.

Davis (1998:13) suggests four reasons why educators should assess:

- To facilitate matching and differentiation

Educators may feel that they cannot present content and activities of an appropriate kind and at an appropriate level unless they know something of their students' cognitive achievements (Davis 1998:13).

- To provide feedback to students on their progress

  Formative assessment is associated with the idea that educators are thereby enabled to provide feedback to students so that they can learn more effectively (Davis 1998:14).

- To enable educators to discover how effective their teaching has been

  If professing educators are so imprudent as to claim that what they are themselves striving to put across is too elusive and too ethereal to be captured by tests or measures of any kind whatever, then they should be told to find something else to teach, something teachable (Flew 1987:35). A sincere intention to teach is simply not possible unless the educator is prepared to assess the results of the attempts to teach (Davis 1998:15).

- To measure the achievements of the educators and the institution

  The results of assessment may be used within institutions to determine the success or otherwise of individual educators (Davis 1998:16). Assessment results may be employed externally by employers as well as by government departments to enable judgement to be made about the quality of institutions. Furthermore, there are continuous drives to use assessment data in a more sophisticated way to determine the extent to which educators and institutions have added value to the students in their drive towards academic excellence.

In conclusion, it can be said that the purpose of assessment could serve to support teaching and learning; to provide information about students, educators and institution; to act as a selection and certifying device; to act as an accountability procedure; and to drive curriculum and teaching (Reddy 2004:33).

Figure 1 provides a schematic summary of the relationship between instruction and assessment (McMillan 1997:7).

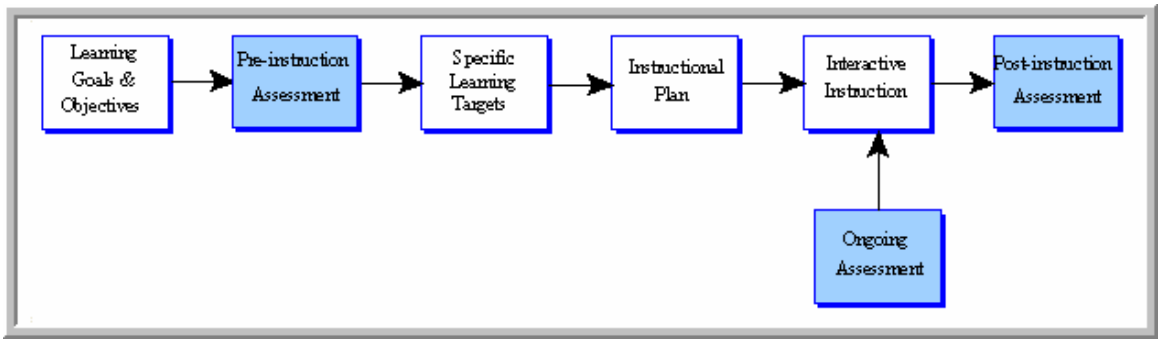


Figure 1 Relationship between instruction and assessment

2.5 APPROACHES TO ASSESSMENT

The terms *assessment* and *testing* are often used by educators and therefore need to be clarified, as this study strongly leans more towards testing, measurement and the interpretation of test scores rather than assessment in its broader sense.

Assessment is a broad and relatively non-restrictive descriptor for the kinds of testing and measuring that educators must do. Assessment is a word that embraces diverse kinds of tests and measurements (Popham 1995:3). McMillan (1997) defines classroom assessment as the collection, interpretation and use of information to help educators make better decisions. Conceptualised in this way, assessment is more than testing or measurement.

There are two broad approaches related to assessment, which are important decision-making factors regarding assessment planning and practice. These two approaches are referred to as summative and formative assessment. It is not the intention of the study to elaborate on the different approaches to assessment, but it is important to note the differences, and to indicate where a CCAT is the most appropriate one to be used.

Summative assessment is a form of assessment that is largely concerned with the final summing up of educational works (Reddy 2004:33). This type of assessment usually concludes a unit of a module or is done at the end of a model, and it is used for making decisions about the students' progress to a next level, or exit points of a qualification. The results from such assessments are judgemental regarding the competency of the students and more to the benefit of the assessor, rather than of the students. Such assessments

overwhelmingly tend to be conducted in the form of, for example, a traditional examination (Bresciani 2004).

After this examination the students receive the results only, with no feedback whatsoever on what has been achieved and what not. It is thus not administered for the sake of enhancing teaching and learning, but only for the sake of grading.

On the other hand, formative assessment is an integral part of learning and takes place on a continuous basis throughout the learning process. Feedback is given a high priority. According to Reddy (2004:34) this form of assessment is seen as being supportive of learning and is non-judgemental, as it focuses on providing constructive criticism. The most important aspect is that it should be capable of determining whether the learning outcomes have been achieved, as this principle is the core of outcomes-based education. It is preferable to provide this information to students on a continuous basis so that it becomes a form of attention and encouragement and an important ingredient of motivation (Reddy 2004:34).

Another important feature of this type of assessment is to inform the educator on a continuous basis of students' progress, and also to address problem areas experienced by students and to adapt teaching strategies in time. The feedback and corrective loop must be as short as possible. One way to keep the loop as short as possible is to be able to make measurements, interpret the scores and provide feedback as effectively and quickly as possible to address problems. With traditional pen and paper tests and the tendency to have larger classes, the feedback loop is not as short as it should be. When feedback is provided after the next section of work is attempted it serves no purpose at all. Students might still have problems with previous learning material, which forms the basis of the next learning material.

From the above it can be seen that there is a definite shift to continuous assessment. Assessment should be formative and ought to contribute to learning. However, it should be kept in mind that it should be feasible to achieve this, without overburdening both the students and the educators. It must be an achievable situation. For this there are some

realities to be considered.

2.6 REALITIES OF ASSESSMENT AT HIGHER EDUCATION INSTITUTIONS

Large classes, research pressure, capacity building, merging politics and a changing funding formula are some of the realities experienced by educators in South Africa (see 1.3).

When considering the above, the introduction of computer-assisted assessment (CAA) becomes a viable alternative for conducting assessment. In order to adhere to most of the requirements of continuous assessment it has become necessary to take a closer look at **how** CAA and testing can be implemented and what the computer technology requirements are.

Henceforth the following two sections will focus on the application and requirements for CAA.

2.7 COMPUTER-ASSISTED ASSESSMENT (CAA)

CAA encompasses the use of computers to deliver, mark and analyse assignments or examinations. It also includes the collation and analysis of data gathered from optical mark readers (Brown et al. 1997).

The use of computers in education is not a new concept. Its first application goes back a long way when it was first used to do psychological testing. It then became clear that it could be applied to more fields in education, especially in the field of testing. Real progress was slow in the early days, as computers were expensive and were only used in large companies. The scenario has changed, however, with the widespread availability of personal computers that enabled educators to focus on the appropriate role of computerisation in the development, administration, scoring and interpretation of tests (Roid 1986:30).

Personal computers have become very powerful in recent years, especially in terms of incorporating high definition graphics for videos and full multimedia without sacrificing speed of operation and flexibility. It is now possible to use these features in computerised testing. Networking together with a Local Area Network (LAN) and the Internet make the deployment and administering of tests much more flexible.

Studies in the field of computer-assisted testing such as those done by Suppes, Atkinson and their colleagues at Stanford (Atkinson & Wilson 1969) are of historical importance because they demonstrated three concepts: (1) that individual student-computer interaction was feasible and cost-effective, (2) that sophisticated hardware and software could be designed for the specialised functions of instruction and testing, and (3) that psychological theories of learning and cognition could be integrated into daily lessons and tests in complex and experimentally meaningful ways (Roid 1986:31).

In the USA, approximately one million examinations for undergraduates and postgraduates were delivered and marked by computers in the 1997-1998 academic year (McKenna & Bull 2000:24). CAA is also a popular assessment method in Australia, with significant usage at the University of Sydney (Dalziel & Gazzard 1999) and at the Curtin University of Technology, which administers approximately 30 000 student tests annually, most of them for summative purposes (Sly & Rennie 1999).

Having said this, the researcher would like to emphasise that although technology shows much promise in contributing to the efforts of test developers, some technical as well as practical problems still remain in developing and implementing computer-assisted testing. The success of CAA lies within the functionality provided by the software. To launch CAA, it is necessary to look more closely at the functionality as provided by the tool as opposed to the specific needs of the educators and the technical feasibility. The requirements for implementing a comprehensive CAA tool are now further discussed.

2.7.1 Using computers to facilitate assessment

A major advantage in using computers as a comprehensive assessment tool is not only the ability to construct and 'bank' test items to subsequently produce a standardised test, but also its ability to administer tests to students and manage student progress records. When the computer is used to administer tests it should be done in such a manner that it exceeds by far the limitations that exist within a traditional pen and paper test, otherwise it will be a vast under-utilisation of computer resources that could otherwise have been used for other purposes. Roid (1986:37) cautions that "computer applications are ill-fated unless they provide a new dimension to testing not possible with non-automated techniques."

Although the role of computers in testing is significant, it is also constrained by three basic conditions (Oosterhof 1994:314):

- There is a greater demand for other applications such as word processing and data management than for classroom testing. This implies that new software products developed for assessment are not readily available on the market.

- The inconsistent nature of classroom activities and computers poses a further challenge. Developers of computer software are at a real disadvantage when addressing the complex needs inherent in the interactions between educators and students.

- Many activities associated with classroom testing are not only unanticipated, but also not well understood. For instance, it is not known with certainty whether improved achievement associated with mastery of learning is the result of integrating assessment and instruction, or simply caused by spending additional time on instruction.

The accessibility of computers in higher education has improved rapidly over the last few years, and as a result the application of technology in the field of education has become more widespread. It must always be considered a tool that assists educators in their task of teaching and it is not intended to substitute the role of the educator.

Oosterhof (1994:315) suggests five applications of computers in the field of education to facilitate assessment. They are:

- development of test items;
- production of tests;
- administration and scoring of tests;
- management of student progress records; and
- assignment of course grades.

If all of the five applications are embedded in a computer program that claims to assist in the field of assessment in education, it could be classified as a comprehensive computerised assessment tool. One can see from the applications to facilitate assessment that all of them centre on a process of testing. In an educational setting assessments are primarily made up of scores obtained from a measurement tool such as a test which is the primary source of obtaining scores. It is thus important to discuss testing as it is part of assessment and forms the core of the comprehensive computerised assessment tool.

2.7.2 Testing

Testing can be defined as the exposure of a person to a particular set of items in order to obtain a score. That score is the end product of testing (Salvia & Ysseldyke 2001:2).

Before any assessment can be made of a student or a group of students, testing must take place in order to obtain a score. It is only from scores obtained by measurements that any assessments can be made thereof. In general terms, testing entails administering a particular set of items to an individual or group of individuals to obtain a score. The score obtained is the end product of testing (Salvia & Ysseldyke 2001:6). Testing may be part of the larger process known as assessment: however, testing and assessment are not synonymous (Salvia & Ysseldyke 2001:6). Assessment in an educational setting is a multifaceted process that involves far more than just administering a test. Testing plays a prominent role in systematic instructional procedures. The main types of items used for testing purposes are objective items (forced choice) and essay (open-ended) items. The

purpose of testing will play a major role in deciding which item type to use, but a CCAT tool must provide the features to handle both item types in its item bank. All features regarding validity, reliability and fairness as well as score interpretation must be applicable to both of the item types where possible. A discussion on the two item types follows below. The study focuses mostly on objective items, since the scores obtained from such tests are reliable.

2.7.2.1 Essay questions

Essay questions refer to questions where students have to provide written answers in the form of calculations, discussions, etc. This CCAT makes provision for the inclusion of essay-type questions in the question bank. Although these questions cannot be marked automatically by the computer, the students can complete the questions online and these questions can then be accessed and marked online by the educator. An example of essay-type questions is displayed in Figure 2.

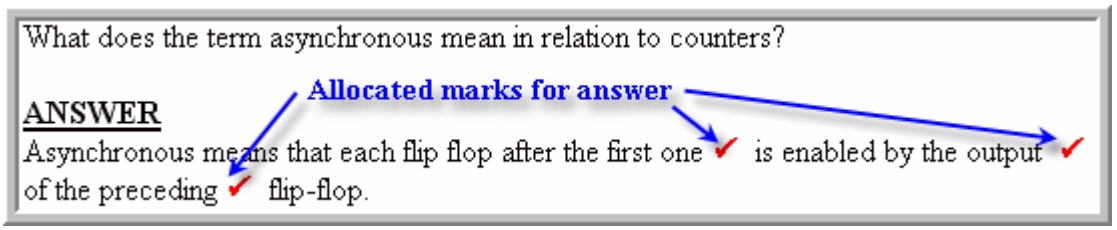


Figure 2 Essay-type item with mark allocation

A very worthwhile feature of the CCAT tool is that the assignment of marks to the answer of an item is stored with the item in the item bank. When setting up a question paper or test, the CCAT will automatically tally these marks as one selects the items to make up the question paper, which is usually a factor to deal with when setting up a test. This will prevent a situation where mark allocation can differ from test to test for the same answer. It will also ensure that mark allocation will be the same even if the test is administered by different educators at different sites. Features included in a CCAT tool are now illustrated using objective items.

2.7.2.2 Objective items

The word *objective* in the term *objective items* refers to the method of marking questions. That is, objective items can be marked without any judgement being made on the part of the marker.

The main purpose of using objective tests is to assess what students know and what they can do (Freeman & Lewis 1998:146). The particular value of this type of testing lies in the fact that students can be tested on a larger domain in a shorter time. This is of particular relevance for reliability since one way of improving the reliability of a test is by increasing the number of items in a test. Concerns that are often raised when assessing a larger domain are that the test will become too long for students to write and that the additional marking load will be too great for the educator to manage properly (Freeman & Lewis 1998). It is the researcher's opinion that objective tests can assess the first three levels of Bloom's hierarchy, namely knowledge, comprehension and application. This view is shared by Sax (1997:103). There are, however, experts who suggest that multiple choice items can be written at any level of Bloom's (1956) taxonomy with the exception of the synthesis level, which requires students to write, present or construct an artefact of their own (Sax 1997:102).

Although research has proved that objective testing can do just as much as essay testing, it is not the researcher's intention to promote objective testing as the only form of testing. It will also become evident that a comprehensive computerised assessment tool incorporates the option of managing essay-type questions (see 2.7.2.1).

Objective items were used in the empirical study for two reasons: reliability and validity indicators are based on the scoring of objective items, and electronic marking had been identified as one of the potential benefits of introducing a CCAT (see 1.9).

2.7.2.3 Objective items with images as distractors

An essential feature of a CCAT is that it must be capable of handling images as distractors.

By asking students to select the correct picture from a list of pictures provided, a higher cognitive level is tested. Students now have more time to spend on reasoning and analysing the circuit instead of battling with primitive drawing tools in their possession to draw the logic circuit, whereas the objective of the item is not to test the students' drawing skills, but rather their understanding of the logic circuit. An example is provided in Figure 3.

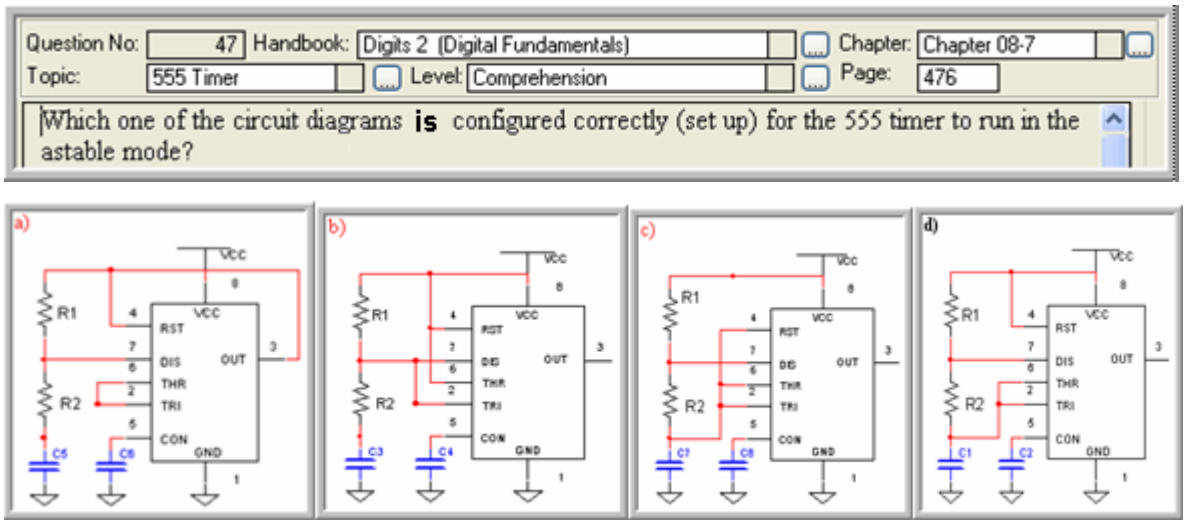


Figure 3 Example of drawings as distractors

2.7.3 Item banking

A test is a quantitative measurement instrument which in turn consists of several items. A test consists of single questions, compiled to form the instrument. These questions are then referred to as **items** of the test (Metsämuuronen 2002:31). According to Metsämuuronen (2002) an item bank is nothing more than a pool of items that are stored after the essential parameters of each item have been stored. Roid (1986:33) states that whether test items are generated by the computer or written offline and simply stored in computer files, it is possible to collect and store them in extensive 'item banks' or 'item pools'. Millan and Arter (1984) define an **item bank** as a "large collection of distinguishable test items". "Large" in this context implies that the bank includes more items than one would use in a single test. "Collection" implies that the items are grouped together not only in a single computer file, but categorised, for example by the textbook from which they have been extracted, the topic they cover and the cognitive level to which they adhere. The grouping of items is shown in Figure 4 where the respective tab headings indicate the different

categories by which the items are grouped together.

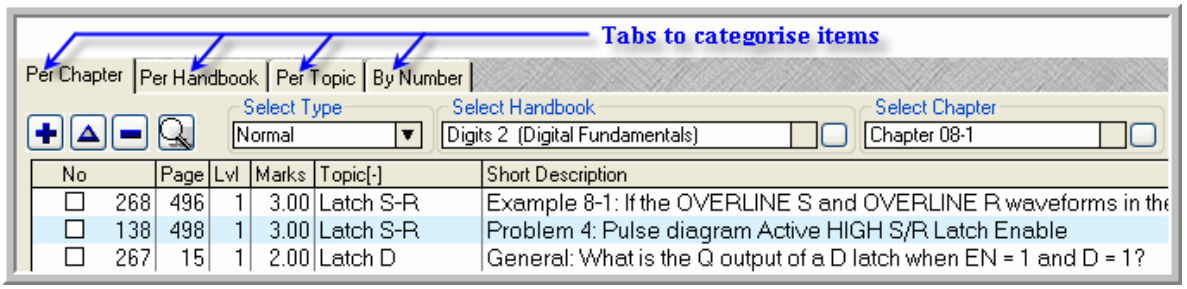


Figure 4 Tabs that categorise items for easy retrieval

For the purposes of this study **item banking** refers to the process of saving a moderated and statistically valid item (question) as part of a larger pool of questions. Metsämuuronen (2002:79) states that the item bank is nothing more than a pool of items. The items are stored after the essential parameters of each item have been found out. It is then possible, when using banked items, to construct balanced tests with pre-known characteristics. With such tests educators will be more capable of determining the real ability level of the students.

Items are written keeping in mind that the main purpose of it is to measure whether an outcome has been achieved. A comprehensive assessment tool should therefore include the option of categorising items according to the outcome it is testing. Moderated items that are banked, together with a functional comprehensive assessment tool, will assist a test constructer in developing balanced, consistent, and content-wise tests, using items of which the difficulty level, discrimination index and other statistical information have been predetermined. It will then be possible to test the real ability level of the students.

All of the aforementioned functionalities and parameters can be built into an assessment tool, but the main concern will always be the **quality** of items, which will remain the responsibility of the subject expert. There is no way that a computer program can evaluate item quality and if the items in a measurement instrument are poor, the assessor will not be able to obtain accurate information (Oosterhof 1994). The measurement results obtained from such a test would then not be worth analysing. The only quality check that the tool can perform is spell checking, which is also limited when items in the natural sciences are

checked. Items are becoming more and more readily available from external sources such as textbook publishers. When importing these items an educator should review the appropriateness and quality of items before inserting them into a question bank (Oosterhof 1994:321). A comprehensive assessment tool should make provision for managing the status of items before accepting them into the item bank. Three statuses are illustrated in the subsequent sections.

2.7.3.1 Item draft status

Figure 5 illustrates items with no marks in the check box, indicating that the item is in a draft mode and cannot be used in a test because it has not yet been checked by a subject expert and/or an assessment committee.

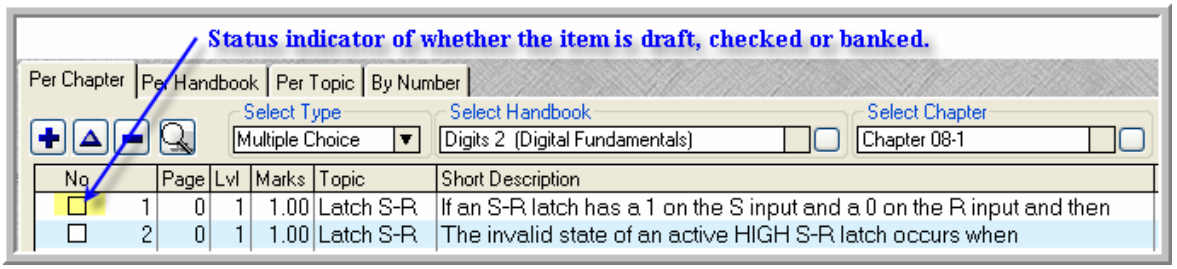


Figure 5 Status of item is 'draft'

2.7.3.2 Item checked status

The tick in Figure 6 indicates that the item has been checked by a subject expert and/or an assessment committee and can now be used for pre-testing or testing. It would be ideal to pre-test every item for banking purposes, but this is not always practically viable because of time and resource constraints. Therefore, if a subject expert, who could also be the educator at a university, is satisfied with an item, it could then be banked without pre-testing. But, as this study stresses, all items used in tests must be analysed, and items that are not up to standard should be pointed out by the analysis.

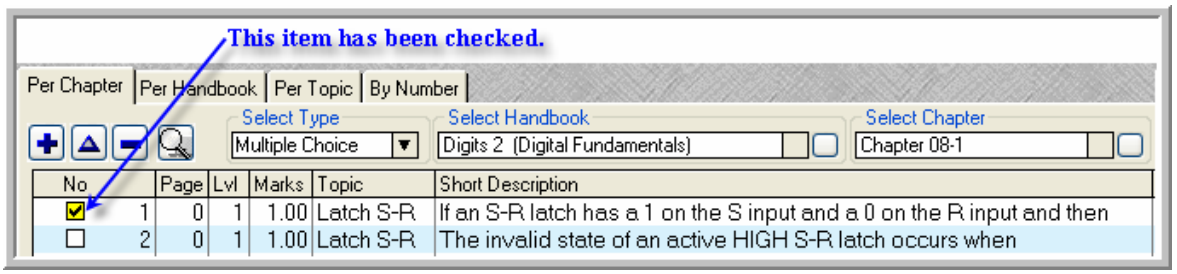


Figure 6 Status of item is 'checked'

2.7.3.3 Item banked status

After pre-testing has been performed on an item, the assessment committee can re-evaluate it by analysing the scores. Any items found to be unsatisfactory must be fixed and, if necessary, retested. If the changes were not significant, the item can be banked if the assessment committee is satisfied, as indicated in Figure 7. The CCAT must therefore make provision for banked items to be revised, but this must only be done by a user with administration rights, thus a user who is password protected.

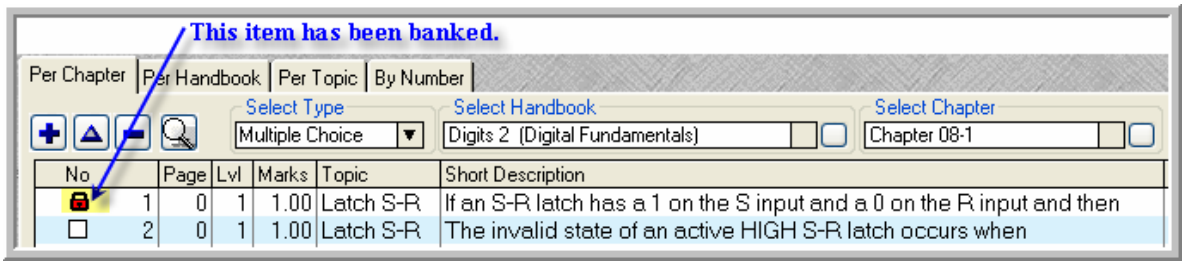


Figure 7 Status of item is 'banked'

The error message in Figure 8 will appear if a user tries to change an item which has been banked.

Figure 8 Status of item is banked

2.7.4 Recording of items

For recording purposes it is important to have information stored regarding the authors and moderators, as well as the date that the status of an item has changed. When an item has been used in a summative event, the moderator's name should be recorded. For terminating modules, the external moderator's name must be recorded. An internal moderator could be a person serving on the assessment committee, such as the assessment officer of the faculty. The caption of moderator information is displayed in Figure 9.

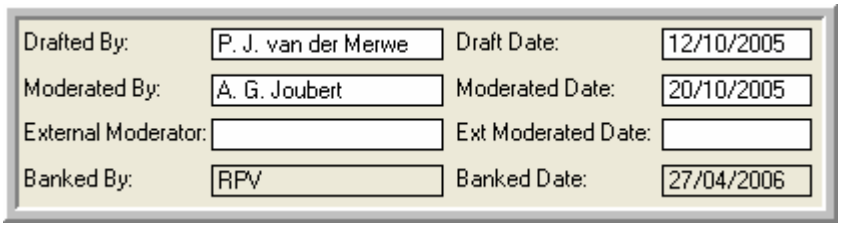


Figure 9 Item quality check information

The check box will be checked and the status will change to a banked item indicated by a red lock as indicated in Figure 7. The 'banked by' field as indicted in Figure 9 will automatically be populated with the name of the user that has logged in and changed the status. This can only be done by a user issued with a high level of security. The date is also captured automatically. After an item has been banked, the status cannot be changed by any user, unless he/she has a higher security level, such as that of the faculty assessment officer.

2.8 STANDARDISED TESTS

A **standardised test** is one that has been carefully constructed by experts in the light of acceptable objectives or purposes. Procedures for administering, scoring, and interpreting scores are specified in detail so that, independent of the person conducting the test or the location, the results should be comparable; and norms or averages for different courses have been pre-determined (Victor 1965:5).

With the increase in student numbers, more educators are required for teaching. When an educator starts to teach in a new environment, he/she often has to make a concerted effort

to grasp the content of a subject, let alone to handle the added responsibility of setting up a test. Standardised tests could be used by all the educators, which in turn will uphold the standard as set by the assessment committee for the subject.

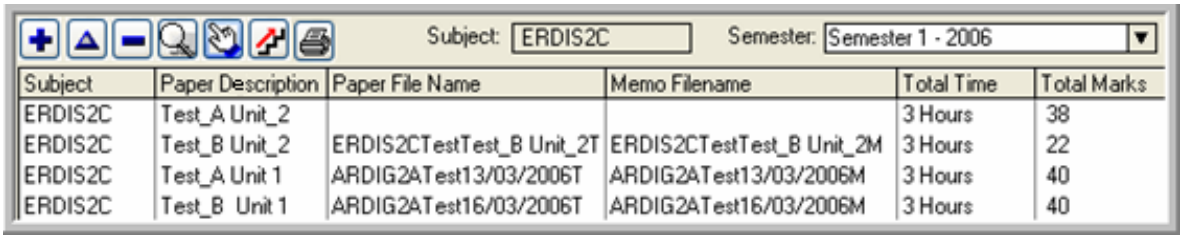
A characteristic of a standardised test that distinguishes it from the educator-made test is that the standardised test has a uniform procedure with respect to administration (Mehrens & Lehman 1969:15). This implies that there are more aspects to consider before a standardised test can be administered. Questions regarding who should administer the test, when the test should be given, and how the test should be administered is very important (Mehrens & Lehman 1969:15). Physical conditions are often neglected by test administrators who ought to make sure that all students take the test under the same physical conditions. Conditions such as seating arrangements, the ventilation, the heat, the lighting and noise levels should always be considered and rectified where necessary.

2.9 ROLE PLAYERS IN THE ADMINISTRATION OF A TEST

An Assessment Committee (AC) should ideally consist of the test administrator, test constructor, test publisher, examiner and a moderator. The examiner and moderator should also be subject matter experts. One person may fulfil more than one role.

Psychological aspects of students taking the test are often neglected. Test anxiety can be greatly alleviated if students are properly informed about the location of the test, test domain, time allocated and number of items in the test. The time of day and fatigue also play a major role when students do tests. To obtain a high level of reliability, all factors that can have a negative impact on the score obtained from the test should be reduced as far as possible. All factors mentioned previously will have a definite impact on the ease with which the test can be administered to students. This then implies that the "test publisher" has an obligation to furnish a manual that provides a description of standard test conditions that must be met. On the other hand, test administrators have an obligation to provide for standard testing conditions and qualified test administrators. Students will only be able to perform maximally – according to clear and concise instructions that both they and the examiner fully comprehend – if these conditions are met (Mehrens & Lehman

1969:19). Figure 10 shows a list of standardised tests of which any one can be selected and administered to students. All the tests for a module are referred to as a *test battery*.



Subject	Paper Description	Paper File Name	Memo Filename	Total Time	Total Marks
ERDIS2C	Test_A Unit_2			3 Hours	38
ERDIS2C	Test_B Unit_2	ERDIS2CTestTest_B Unit_2T	ERDIS2CTestTest_B Unit_2M	3 Hours	22
ERDIS2C	Test_A Unit 1	ARDIG2ATest13/03/2006T	ARDIG2ATest13/03/2006M	3 Hours	40
ERDIS2C	Test_B Unit 1	ARDIG2ATest16/03/2006T	ARDIG2ATest16/03/2006M	3 Hours	40

Figure 10 An example of a test battery compiled for a module

To assist in setting up a test that will be well balanced across the cognitive levels, a specification grid must be part of the assessment tool.

2.10 SPECIFICATION GRID

Imrie (1995) advocates the use of an "assessment planning grid" underpinned by a taxonomic as an "indicator of quality assurance". Similarly, Heard, Nicol and Heath (1997) suggest the completion of a test specification matrix, which measures learning levels (based on Bloom's taxonomy), against the topics covered in the assessment. If such a test specification grid is used, the test designer can analyse the breadth and the levels of learning assessed by the examination and correct for areas of poor coverage. Additionally, external moderators can use such information in judging the suitability of the assessment and its relationship to the National Qualifications Framework (NQF) and Bloom's Taxonomy prescribed for the course/module objectives on which the OBE system is largely based. This implies that a test cannot be compiled manually by selecting items, or when the tool automatically selects items randomly, without taking the specification grid into consideration. Figure 11 shows an example of a specification grid set up for a specific module, in this case Logic Design. The percentage allocated to each cognitive level will depend on the NQF level within a course in which it is presented.

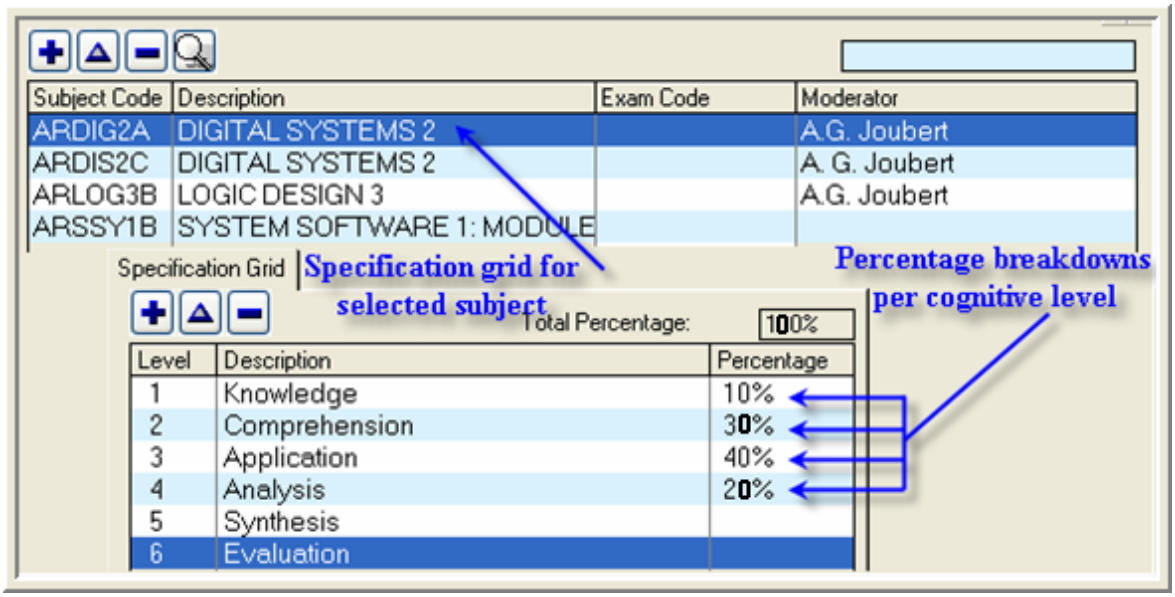


Figure 11 An example of a specification grid

When an item is entered into the CCAT the cognitive level of the item must also be entered, as indicated in Figure 12.

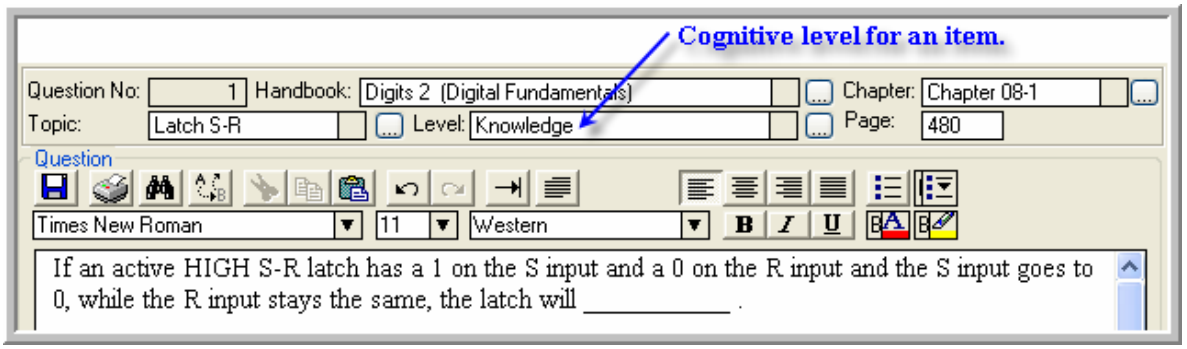


Figure 12 Bloom's taxonomy entered as part of the parameters of an item

This will in turn reflect when a test is generated, and it will compare it to the specification grid for that module which is referred to in CCAT as the complexity distribution of a particular test. The complexity distribution example in Figure 13 shows under the *Subj %* column that the percentage knowledge items are too low, comprehension is also too low and application questions too high. By sticking to the percentages as set up by the specification grid it is much easier to prepare a well balanced question paper, which will in turn conform to what is expected by the NQF.

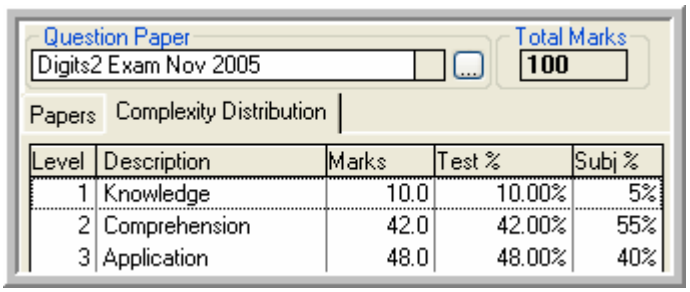


Figure 13 Specification grid for a question paper

2.11 ITEM REPETITION CHECK

It is of utmost important to Engineering Council of South Africa (ECSA) that items are not repeated at least within the two previous examinations. It is required of staff of the Faculty of Engineering to submit the last two examination papers to the moderator who will check that items are not repeated. This can become a burden to educators in setting the examination paper if they have to page through the previous examination papers manually to check for unnecessary duplication. The CCAT tool has the feature that it logs the semester in which the item was asked in an examination as well as a test. This feature will ensure unnecessary duplication of items in examinations as well as in tests. Figure 14 provides an example of an item on the left-hand side and the corresponding tests in which the item appears on the right-hand side.

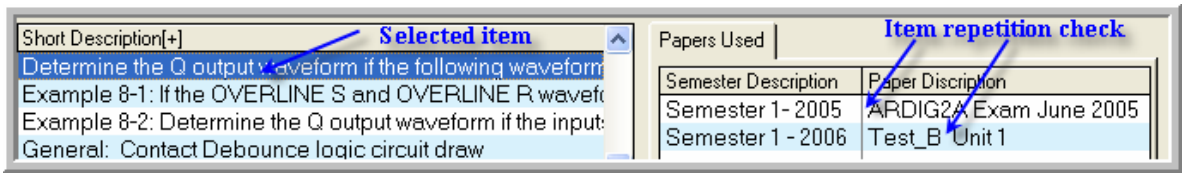


Figure 14 Item repetition check

2.12 ADMINISTERING AND SCORING OF TESTS

The main purpose of administering tests to students is to obtain and interpret scores. This score is very important in the sense that important decisions regarding the educational process are made from the scores. This implies that a proper analysis of the score must be done. Educational measurement has the potential to develop numerous and extensive item banks at regional and university centres at a national level. Recent years have seen a

proliferation of published criterion-referenced achievement tests that feature the possibility for school districts to adopt test content to their particular curricular emphasis. At VUT, where this study is conducted, such item banks must be developed for all the modules presented in a department. These developments on item banking and standardised tests are part of the movement towards a closer linking of testing and instruction in educational program evaluation (Roid 1989:33).

Another arena in which educational measurement has broken new ground is in the development of theories and statistical models. Perhaps more than any other contribution of computer technology to testing, the use of large-scale computers in the sophisticated, multivariate analysis of test data has made a major contribution to the overall improvement in the precision and accuracy of educational tests during the past century (Roid 1989:33). Due to the importance of tests in determining the future of a student, and the requirements set by the governing higher education bodies (HEQC, ECSA) for quantitative indexes of reliability and validity, it is difficult to imagine a test that could be developed and administered to students without any form of computerised analysis.

The reality of working with human beings imposes factors such as students not showing up for a test, for various reasons varying from acceptable to not acceptable. But the reality is that the situation needs to be accommodated. This leads to the need for additional data-handling features to be incorporated in the assessment tool (such as a flag indicating absence from a test). A facility to ignore test results for a specific student when calculating grading is also an important feature.

Note fields as part of the database at various places within the assessment tool enable the educator to enter notes for future reference. The program must make provision to manually enter scores obtained from pen and paper tests individually or as part of CAA.

In Figures 15(a) and (b), two screens are displayed to indicate the user-friendly interface for educators to assist in administration.

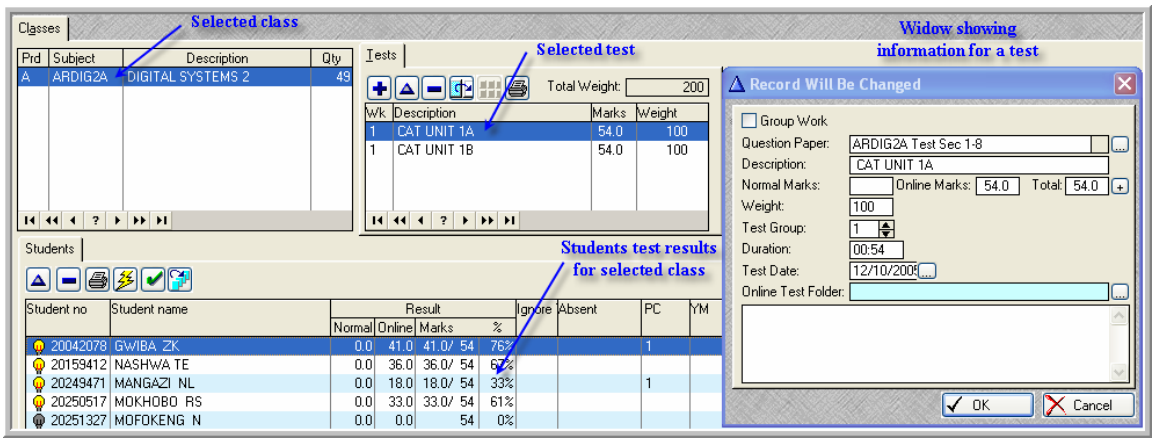
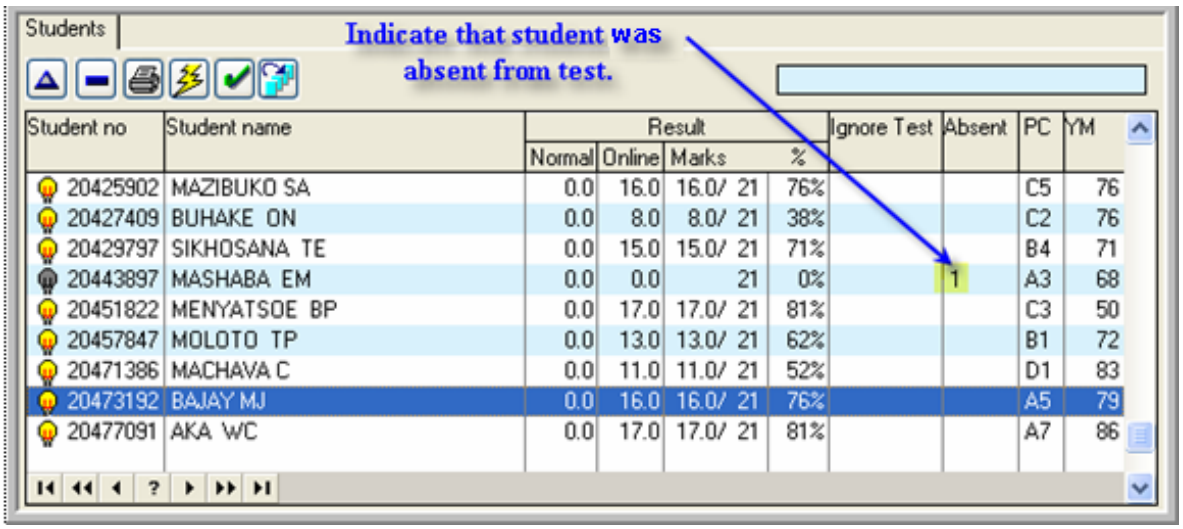


Figure 15(a) Test information window



Student no	Student name	Result				Ignore Test	Absent	PC	YM
		Normal	Online	Marks	%				
20425902	MAZIBUKO SA	0.0	16.0	16.0/ 21	76%			C5	76
20427409	BUHAKE ON	0.0	8.0	8.0/ 21	38%			C2	76
20429797	SIKHOSANA TE	0.0	15.0	15.0/ 21	71%			B4	71
20443897	MASHABA EM	0.0	0.0	21	0%		1	A3	68
20451822	MENYATSOE BP	0.0	17.0	17.0/ 21	81%			C3	50
20457847	MOLOTO TP	0.0	13.0	13.0/ 21	62%			B1	72
20471386	MACHAVA C	0.0	11.0	11.0/ 21	52%			D1	83
20473192	BAJAY MJ	0.0	16.0	16.0/ 21	76%			A5	79
20477091	AKA WC	0.0	17.0	17.0/ 21	81%			A7	86

Figure 15(b) Educator's interface showing different tables for administering test scores

2.12.1 Management of student progress records

Within education, computer management of records makes it possible to track students, to determine promptly what each student has achieved, and to formulate plans for future instruction (Oosterhof 1994:324). The data management capabilities of computers can facilitate the integration of testing and instruction and thus help personalise learning. Oosterhof (1994) listed several key factors that influence the effectiveness of data management programs. These factors will be discussed, keeping in mind what is specifically needed in an educational environment and how it is accommodated or how it is lacking in the different assessment tools.

- The data-management system must retain and be able to retrieve useful information.
- A data-management system must be able to summarise information.
- A data-management system should anticipate the needs of the educator.
- A data-management system should be easy to learn.
- A data-management system must be easy and efficient to use.
- A data-management system must provide valid information.
- When possible, the data-management system should automate the management of students.

2.12.2 Assignment of course grades

Many educators soon discover that maintaining student records and later assigning end-of-term grades are not enjoyable tasks. The worst part of these chore is the weighting and totalling of each student's score (Oosterhof 1994:327). With the prospect of doing continuous assessment with various scored formative events this is a major consideration in handling all of the scores of each event for each student. Not only are there numerous scores, but at the same time there are many student scores that have to be interpreted.

Using a computer-assisted testing tool can be a great aid to the educator in setting up items, easily managing them and administering tests to students effortlessly and accurately. However, the next critical concern after testing is the administration of the students' results. The requirements of strategies for continuous assessment necessitate more testing, which more often leads to more results. Students' results must at all times be recorded and kept. Hard copies of results can be made, but there are many more advantages to having marks electronically available. The ability to analyse such marks by using statistical analysis and graphs makes the scores much more meaningful.

Applying all or most of the continuous assessment strategies in a module leads to the requirement of a CCAT to be able to assign weights to the different tests. Different criteria can be used in deciding on the weight that must be assigned to each assessment. In a continuous assessment environment it is important to decide how much each assessment or test will count in the final grade (McMillan 1997:309). This is important for calculating a

student's year mark which contributes 50% toward the final examination mark. Obviously, more important assessments or tests will be assigned higher weights. The question now is: What determines whether an assessment is important? The answer is not straightforward, but the educator's professional judgement is always important. According to McMillan (1997), the educator can be guided by the most significant assessments or tests as those that:

- correspond most closely to the learning goals and targets (content-related evidence for validity),
- reflect instructional time,
- are most reliable, and
- are most current.

Most modules are structured in such a way that three marks in total must be submitted to the ITS (institutional data-recording system used at VUT) for a module obtained during the semester. The need now arises from this to be able to group any amount of tests together, which should add up to a single mark with the corresponding weight of every test taken into consideration.

By setting the weight to zero for a particular test it should ignore the entire test so that it will not contribute to any of the test groups. A flag can also be set for an individual student to ignore the test result, so that when calculating the earmark it will not consider the test result for that student. An absent flag can be set to inform the educator that the student did not write the test so that a zero result cannot be confused with an absent that will also produce a zero result. All the results are stored under the semester date in which they were obtained. As results are entered into the system and the year mark is calculated for what is done so far. This supports the idea of continuous assessment as a student's progress is monitored in an ongoing fashion.

A backup file of all these results is made from the backup option within CCAT. It can be burned onto a CD and kept in the administration office for future enquiries. Maintaining and administering test results must be made as easy as possible for the already overburden educator.

2.12.3 Register

Whether it is important to know of the absenteeism of students or not is debatable. However, a CCAT should include a feature whereby a student can be marked absent. The feature can be used or not and will not have an overall effect on the tool. Figure 16 shows the absence of students for a specific subject and the time that they were absent. This is indicated by a 1.

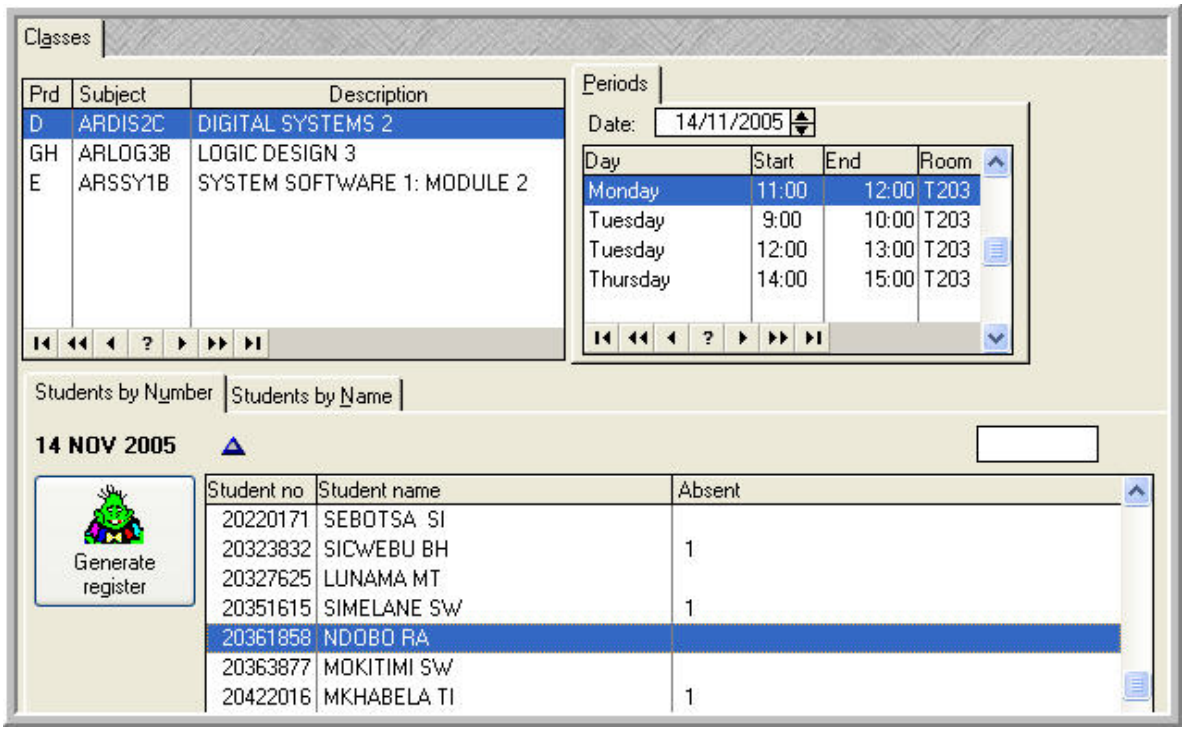


Figure 16 Electronic register

2.12.4 Group/team work

Group work is a very important facet of outcomes-based education. A CCAT must, therefore, also be capable of handling the record keeping of the students within the group, together with the scores obtained by the different groups. Figure 17 shows an example of students within a class that are assigned to different groups.

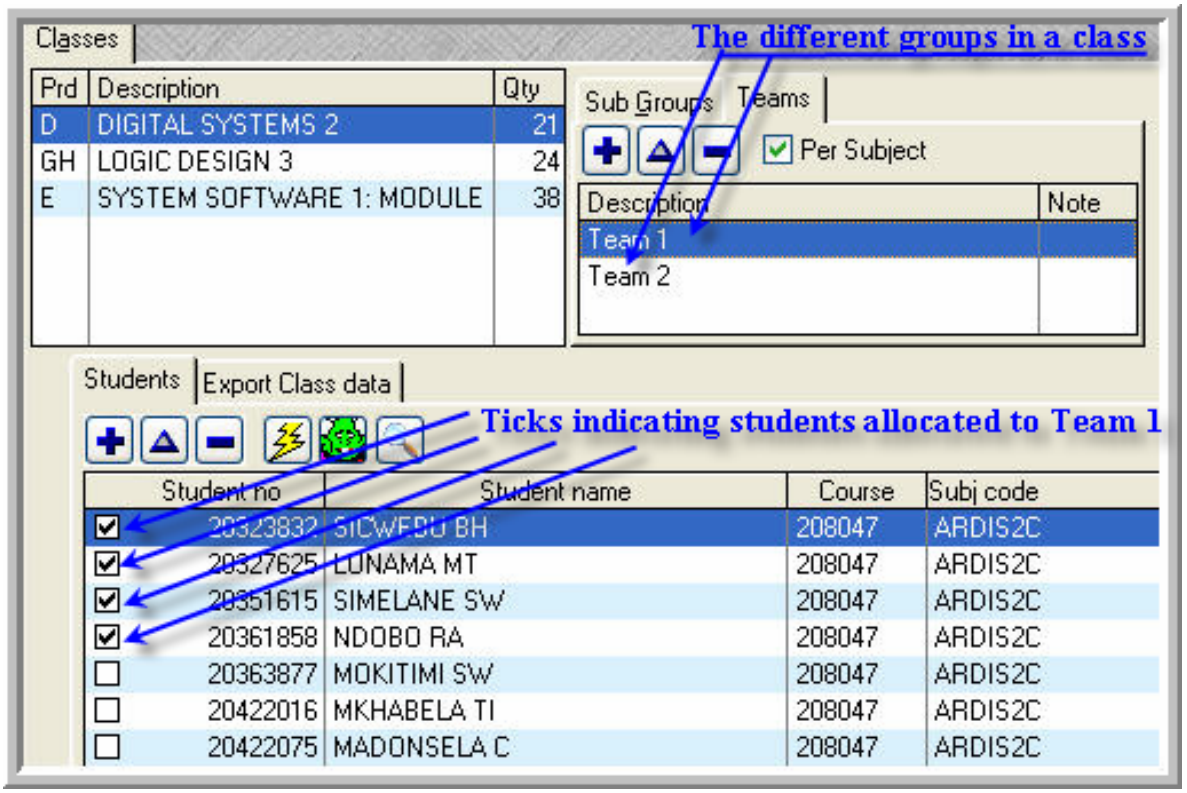


Figure 17 Administering group work

2.12.5 Reporting

The reports, as indicated in Figure 18, can be generated from the CCAT tool. Flexibility is provided as to the preferred format of the output.

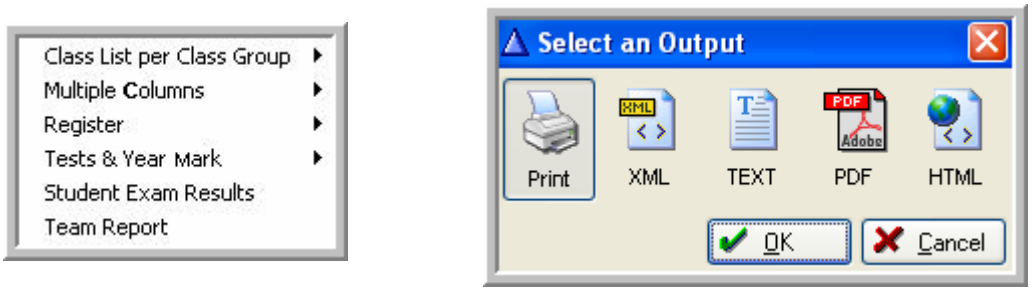
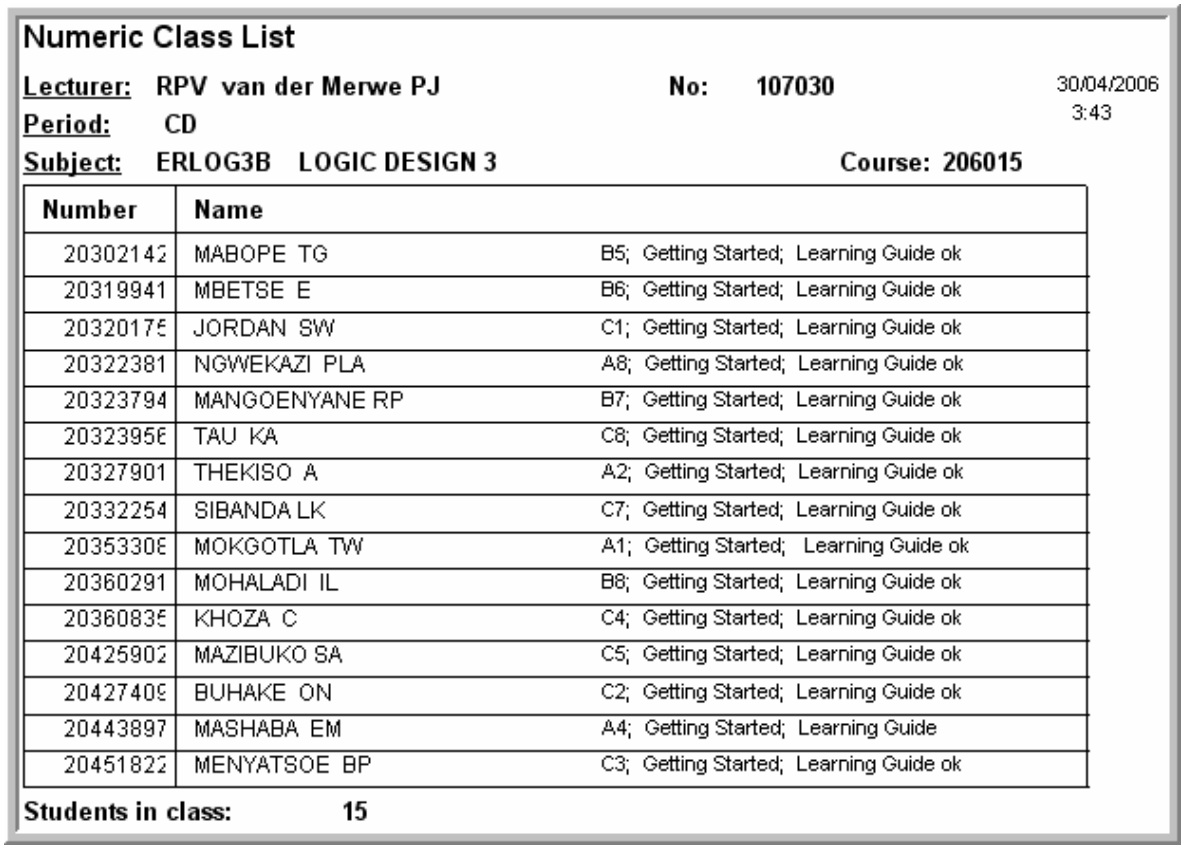


Figure 18 Report generating facility

Figure 19 illustrates a typical report, containing student records as well as additional information, for example, whether they have received a learning guide and their respective seat numbers in the laboratory.



Numeric Class List

Lecturer: RPV van der Merwe PJ **No:** 107030 30/04/2006 3:43

Period: CD

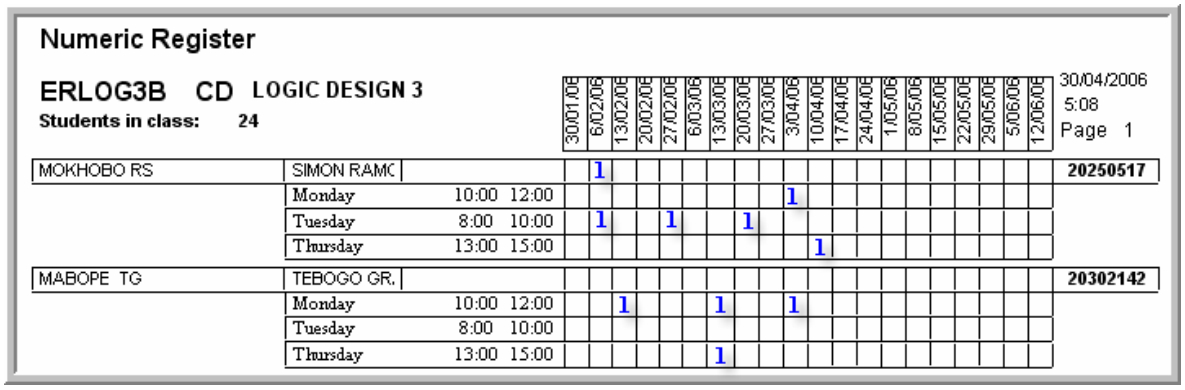
Subject: ERLOG3B LOGIC DESIGN 3 **Course:** 206015

Number	Name	
20302142	MABOPE TG	B5; Getting Started; Learning Guide ok
20319941	MBETSE E	B6; Getting Started; Learning Guide ok
20320175	JORDAN SW	C1; Getting Started; Learning Guide ok
20322381	NGWEKAZI PLA	A8; Getting Started; Learning Guide ok
20323794	MANGOENYANE RP	B7; Getting Started; Learning Guide ok
20323958	TAU KA	C8; Getting Started; Learning Guide ok
20327901	THEKISO A	A2; Getting Started; Learning Guide ok
20332254	SIBANDA LK	C7; Getting Started; Learning Guide ok
20353308	MOKGOTLA TW	A1; Getting Started; Learning Guide ok
20360291	MOHALADI IL	B8; Getting Started; Learning Guide ok
20360835	KHOZA C	C4; Getting Started; Learning Guide ok
20425902	MAZIBUKO SA	C5; Getting Started; Learning Guide ok
20427409	BUHAKE ON	C2; Getting Started; Learning Guide ok
20443897	MASHABA EM	A4; Getting Started; Learning Guide
20451822	MENYATSOE BP	C3; Getting Started; Learning Guide ok

Students in class: 15

Figure 19 Class list report

An example of a hard copy of the register of absenteeism is shown in Figure 20



Numeric Register

ERLOG3B CD LOGIC DESIGN 3

Students in class: 24

30/04/2006 5:08 Page 1

			30/01/06	6/02/06	13/02/06	20/02/06	27/02/06	6/03/06	13/03/06	20/03/06	27/03/06	3/04/06	10/04/06	17/04/06	24/04/06	1/05/06	8/05/06	15/05/06	22/05/06	29/05/06	5/06/06	12/06/06	
MOKHOBO RS	SIMON RAMC			1																			20250517
	Monday	10:00 12:00										1											
	Tuesday	8:00 10:00		1			1			1													
	Thursday	13:00 15:00											1										
MABOPE TG	TEBOGO GR.																						20302142
	Monday	10:00 12:00			1				1			1											
	Tuesday	8:00 10:00																					
	Thursday	13:00 15:00							1														

Figure 20 Register report

2.13 CONCLUSION

This chapter explored the role that computers can play in the assessment process and analyses the spectrum of activities which comprise computer-assisted assessment. The chapter also reflected on the design and development of a CCAT by indicating how the functionality is implemented and accommodated in the tool.