CHAPTER 1

INTRODUCTION AND PURPOSE OF THE STUDY

1.1 INTRODUCTION

For many years tests have been administered to students. The main reason for letting students write tests is to obtain a measurement of what a student has learned of a certain domain. Wither (1994:13) defines **assessment** as the process of making a decision about something, of examining or testing the performance of individuals, and of evaluating or judging on the basis of criteria such as correctness, validity and empirical evidence. The process and outcome of assessing would therefore be an assessment.

The educational achievement must be measurable. However, the result of the measurement has far-reaching effects on educational progress as a whole. The reason is that the results of the measurements are used to evaluate the educational process (system) as a whole, which includes the institution as well as the educator. In fact, it also determines the final output of the institution, which has an influence on the government subsidy. Most important of all, it is used to grade a student, and to decide whether the student will be promoted to the next level of study. At the end of a programme, the results of the measurements will determine whether or not a student is ready to enter the workplace as a competent individual.

The introduction of outcomes-based education and training has required a different approach to education, including assessment (Department of Education 1997:3). An outcomes-based approach to education and training focuses on continuous assessment through the use of a range of assessment methods (Sutherland & Peckham 1998:100).

The abovementioned scenario, coupled with the increasing use, affordability and flexibility of technology, clearly necessitates a closer look at the application of technology in assessment. According to O'Reilly and Morgan (1999) the move to online and computer-

assisted assessment is a natural outcome of the increasing use of information and communication technologies to enhance learning.

In the USA, approximately one million examinations for undergraduates and post-graduates were delivered and marked by computers in the 1998-1999 academic year (Bennet 1998; McKenna & Bull 2000:24). Curtin University of Technology in Australia administers approximately 30 000 student tests annually (see 2.5), most of them for summative use (Sly & Rennie 1999). Higher education in the UK has also taken a strategic decision to embark on the road of computer-assisted assessment (Mckenna & Bull 2000:24).

The increase in student numbers, often coupled with a decrease in staff resources, motivated higher education institutions in South Africa to search for more effective ways of assessing large student numbers. Between 2000 and 2003 student enrolments grew by 22 percent while teaching and research staff grew by only six percent (Blaine 2005:5).

1.2 OBJECTIVE OF THE STUDY

With few exceptions, much of the development and integration of technology and assessment have been done in an ad hoc fashion. As the uptake of computer-assisted assessment grows, the need for a comprehensive assessment instrument becomes critical. This instrument should be able to take measurements and provide scores and statistical analysis in an understandable format so that proper interpretations of the results can be made. If measurements of tests are acquired and cannot be properly evaluated, the results of the measurements are useless. In such a case it can be argued that the test as a whole is of no value at all. If the reliability cannot be determined, the validity of the whole assessment process can be questioned (Dessus, Lemaire & Verner 2000; Gardner, Sheridan & White 2002).

Black and Harrison (2000) report on research results that positively indicated that formative assessment practices on a continuous basis produced significant learning gains. It is not a viable option to evaluate measurements manually. The literature study will show

that test results are complicated when all aspects (statistical) of measurements are included. A computerised assessment tool is required that will not only do the measurement, but will also assist the facilitator in administering tests to students.

This study aims to investigate the validity, reliability and fairness of computer-assisted measurements by means of online statistical analysis. In order to assess the quality of the items that make up a test, it is important to know the essential parameters of the items. The parameters of items are also important when banking items (Metsämuuronen 2002:32). The aim of this research was to develop a measurement instrument that will assist in test construction that will give us a more truthful picture of the measurement obtained from computer-assisted assessment.

This measurement instrument was implemented as an integral component of the Comprehensive Computer-Assisted Assessment Tool (CCAT) used to conduct the study.

1.3 TERM CLARIFICATION

It is necessary to conceptualise the following terms as related to the study:

Computer-Assisted Assessment (CAA)

Computer-assisted assessment encompasses the use of computers to deliver, mark and analyse assignments or tests (Brown, Bull & Pendlebury 1997).

Validity

Deale (1975) defines validity as the quality which a test should have if it is to achieve the outcomes that are intended. According to Salvia and Ysseldyke (2001:679), a test can be considered as valid if the test indeed measures what the assessor claims it measures. Test validity also concerns the appropriateness of the inferences that can be made on the basis of test results. Validity refers to the appropriateness, meaningfulness and usefulness of the specific inferences (Alberto & Troutman 1990).

Reliability

Deale (1975) defines reliability as consistency, meaning how far the test would give the same results if it could be done again by the same students under the same conditions. In measurement, reliability refers to the extent to which it is possible to generalise from an observation of a specific behaviour observed at a specific time by a specific person to observations conducted on similar behaviour, at different times, or by different observers (Salvia & Ysseldyke 2001:679).

Fairness

Assessment arrangements should not advantage any one student over another – thus all assessment opportunities should be equal. Fairness also refers to the marking process that should be unbiased (Freeman & Lewis 1998:306).

1.4 PROBLEM STATEMENT

Assessments are administered to students by individual educators and also set up their own tests, sampling a domain. This often results in the following problems:

- The complexity and time-consuming effort needed to do a proper evaluation of the results are not done by the educators. This evaluation is necessary because it states whether a test is valid and reliable.
- All educators are not necessarily good item writers. This does not imply that they are not good facilitators. To write good items, educators need special skills and training to become good item writers.
- With student numbers increasing per class, it is perceived that the minimum tests are written. This is not in line with the principles that underpin continuous assessment. More assessments are required to get more and quick feedback to students. Due to limited staff and time resources, the assessment process is not flexible (Brown *et al*. 1997).
- Different educators administer different items; therefore sampling of the domain might not be the same. Most often it is not the same. Biasing towards sampling of the domain might occur, for example sample only that which has been taught, or that which has been

best understood by the students. This is done to project a good test result.

- When using essay-type items the marking may not be reliable, although the items might be valid (Freeman & Lewis 1998).
- When educator-made tests are administered to students it is not possible to highlight shortcomings on the facilitator's presentation of a module (Freeman & Lewis 1998:307).

1.5 RESEARCH QUESTIONS

Based on the above discussion the following questions arise:

- What characteristics comprise an effective Comprehensive Computer-Assisted Assessment Tool (CCAT)?
- What are the operational requirements necessary for the implementation of the Comprehensive Computer-Assisted Assessment Tool (CCAT)?
- Do the results of computerised statistical analysis reflect the validity, reliability and fairness of the measurements?

1.6 RESEARCH OBJECTIVES

In an attempt to answer the above-mentioned questions, the following objectives for the research project were identified:

- Determine the technical, pedagogical and operational requirements of an effective Comprehensive Computer-Assisted Assessment Tool (CCAT).
- Determine what values are required to incorporate a good quantitative measurement tool.
- Develop a tool that will assist facilitators in giving a reliable, valid and fair scoring of the measurement of their students' performance.

1.7 DELIMITATIONS OF THE STUDY

The study will focus primarily on the statistical evidence that can be acquired by objective types of tests. Essay-type test cannot be evaluated to a satisfactory degree with statistics because they cannot be marked by machine.

Although CCAT can be implemented in any discipline, this study will focus on the implementation of CCAT for formative assessment (see 2.3) of Digital System students' tests.

1.8 RESEARCH METHODOLOGY

1.8.1 Literature study

A comprehensive literature study on tests and the measurement thereof, as well as on tools that are currently available, was conducted by consulting journals, the Internet, textbooks and databases (Nexus, Sabinet, Ebscohost, SwetsWise and Emerald).
The keywords used in the searches are validity, reliability, fairness and computer-assisted assessment tool.

1.8.2 Empirical research

The research was quantitative in nature. CCAT was implemented as an assessment tool. Results of CCAT formative assessment events were statistically analysed to reflect on the validity, reliability and fairness of the measurements.

1.8.2.1 Population

All students enrolled for Digital Systems at the VUT during the second semester of 2005 wrote a pre-scheduled number of formative assessment tests using CCAT.

1.8.2.2 Procedure of data collection and analysis

- Regarding the ethical aspects and other possible issues of the research, tests were not administered to students unless written consent was obtained from the Head of Department of Computer Systems for permission to use the department's students and subject content, and the students themselves for taking part in the research project as volunteers.

A letter of permission was also obtained from the Director of Technology to allow the use of technology facilities and technical support for the period required for the purpose of the study.

The data was collected over a period of 12 weeks.

1.8.2.3 Data analysis

To establish what the item parameters were and how they measured the quality of an item, it was necessary to collect various quantities of data. These quantities could only be obtained from items constructed and then answered by the students. The data collected were used to enhance the tool which in turn would then determine the quality of the item, and make it suitable for banking. The quantitative data collected would assist in determining the following statistical values:

- Item difficulty;
- Item discrimination;
- Distractor analysis;
- Kuder Richardson (KR_{20});
- Kuder-Richardson (KR_{21});
- Pearson product-moment correlation coefficient;
- Content validity;
- Cronbach's Alpha;
- Spearman-Brown prophecy.

1.9 IMPORTANCE OF THE STUDY

Measurements take place in almost every sphere of life. We as individuals are constantly being measured, for example by our managers, peers and children. We are also constantly measuring the outcomes of sports events, politics, or a show we attended. Measurement is evident in all disciplines. It is of paramount importance to education and learning.

Learning is a process of being measured, and then most importantly, of being given feedback. Without feedback learning cannot take place. The more measurements there are and feedback given, the better learning should take place. This is the basis of continuous assessment. The following significant benefits could result from this study:

A Comprehensive Computer-Assisted Assessment Tool (CCAT) to support continuous assessment could be implemented in other disciplines as well, as it will be easier to administer tests to students.

Marking overloads could be reduced, as most of the formal formative assessments could be marked by a CCAT. Large student numbers will not have a major impact on facilitators' workload.

Administering objective tests will always yield a hundred percent reliability as well as validity if items from the item bank are used. When essay-type questions are used from the item bank they should have a high validity value, although reliability may vary according to the marking of the scorers.

Institutions of higher education (HE) are held accountable for their success or failure in the promotion of learning. If measurements of tests are found to be valid, reliable and fair, then a more accurate evaluation of the module could be determined. Problems could be detected and rectified in time. It can then be more effectively proven to the qualification accreditation bodies such as the Engineering Council of South Africa (ECSA) that all measurement were done by analysing the results with proven statistical measures.

To obtain measurements of all students, irrespective of their facilitator, the tool will make provision for item banking as well as for standardised tests. This will result in tests being administered to students using only moderated items. Only moderated items can and will be allowed by the tool to be banked.

The problem of administering tests at multiple sites of delivery will be alleviated greatly due to the fact that the facilitators can also use items from the item bank. Standardised tests

will also be administered more easily from the main campus.

Due to the ease with which standardised tests can be compiled by the examiner, re-testing can be easily achieved. This will allow for a more flexible assessment system.

1.10 OUTLINE OF PROPOSED STUDY

Chapter 1: Introduction and purpose of the study.

Chapter 2: Requirements of an effective Comprehensive Computer-Assisted Assessment Tool. This chapter focuses on the technical, operational and pedagogical requirements, and provides a critical overview of quantitative measurement values required to ensure validity, reliability and fairness of items included in a question bank.

Chapter 3: Research design. The different phases of the research process are outlined and described.

Chapter 4: Reliability statistical analysis. The objective of this chapter is to report on the results of the statistical analysis of the measurements to reach a conclusion regarding the reliability of items.

Chapter 5: Validity statistical analysis. This chapter reports on the results of the validity analysis of the measurements to reach a conclusion regarding the strategy of item banking and the construction of standardised computer-assisted tests.

Chapter 6: Conclusion and recommendations. In this chapter the findings as formulated in this dissertation are concluded and a recapitulation of the findings of the complete research project are provided.