# CHAPTER 9

## ESTABLISHING A PROCESS OF DIAGNOSTIC ASSESSMENT

### 9.1 The final intervention and action research cycle

Chapters 5 and 6 highlighted three phases in a reading intervention, considered the reasons for their introduction, and the results they achieved. Three action research cycles were involved in the analysis of these three phases. Chapter 7 describes three phases of the use of an alternative approach to formative assessment. The results were analysed and modifications were made, again involving three action research cycles. These two parallel sets of interventions all had a common aim: to improve the success of students taking the Mathematics Access Module (success in terms of passing the examination, so that they could move on to study mathematics at a higher level).

From the results discussed in Chapter 8 we see that although there were fluctuations in the pass rate none of the activities undertaken with time to improve the way the module was taught, and none of the interventions, separately or together, had a significant impact. The final cycle considered in this thesis is a consequence of the overall performance of students, and the research results discussed in the previous chapters.

In this chapter we consider the rationale for diagnostic assessment, i.e. the need to identify in some way students at risk of failing the Mathematics Access Module. Related to this is the tension inherent in testing within a distance learning context. We consider briefly entry-level testing from an international and a South African perspective. Chapter 9 then discusses UNISA’s acceptance of diagnostic assessment for Mathematics Access Module students, the feasibility of introducing assessment, assessment format and cost. It focuses on an investigation into two possible options, namely the University of Cape Town Alternative Admissions Research Project (AARP) tests, and the placement test battery used at the (then) University of Port Elizabeth (UPE). One of the UPE tests that seemed appropriate for UNISA, for a variety of reasons, was the ACCUPLACER Reading Comprehension (ARC). Chapter 9 discusses the reasons for the selection of the ARC and the development of a UNISA-specific test for quantitative reasoning (the Basic Arithmetic Test (BAT)) as the two components of the Mathematics Diagnostic Assessment (MDA) implemented in 2004. The components of the ARC and BAT are described, and the rationale is given for the scoring procedures used. These scores were consolidated into three risk categories, which are then discussed. The chapter gives the initial results obtained in the

MDA in January 2004; also later results once the MDA was written for a second time, in June 2004. The comparison of MDA and Mathematics Access Module assignment results is given. Chapter 9 ends with a discussion regarding possible support options that could arise from an identification of risk potential.

The research described in this thesis ended before the 2004 examinations were written. The preliminary analysis of MDA results suggests that the diagnostic assessment process could be further modified; results need be further analysed. There is clearly scope for further research in this regard.

### 9.2 Rationale for the introduction of diagnostic assessment at UNISA

Over the years of its existence, the Mathematics Access Module has attracted increasingly large numbers of students, with limited levels of success. The previous chapters highlight some of the factors that have contributed to their problems. Many students have not been able to pass the Access Module even after more than one attempt; many students have not been able to benefit from the assignment process, and have not been able to obtain admission to the examinations. Chapter 5 (investigation of reading skills and face-to-face reading intervention), Chapter 6 (video intervention and further investigation of reading difficulties) and Chapter 7 (alternative assessment) have dealt with the various initiatives that were introduced in an attempt to promote success for Mathematics Access Module students. Up to the end of 2002 none of these initiatives appeared to be having any measurable impact on student results. There was, however, a considerable improvement in the pass rate at the end of 2003 (see Table 8.9 in Chapter 8). From 14,9% in 2002 it increased in 2003 to 27,0%. 2003 was the first year in which there was some sort of a ‘test’ which would have given students early feedback on potential reading problems. It was undertaken by most students as part of the first assignment. The results of this ‘diagnostic test’ (see Chapter 6) supported the introduction of a more generalised and structured assessment.

There are many prerequisites for successfully studying the Mathematics Access Module, namely the existence of a certain minimum level of numerical and mathematical skill, the willingness and determination to cover a wide range of concepts, and competence in reading and communicating in English, the language of instruction. It has become increasingly evident that many Access Module students tend to underestimate these prerequisites. Students who attempt tertiary-level study with insufficient levels of skill or background need more support than can be provided in the module itself. The prediction made by Lewin (1997) (see 1.4.2 in Chapter 1) seems to have been valid: the quality of students would change a result of massification, thus leading towards a situation where incoming students would require a wide range of learning support measures. Lewin (1997) also suggested that at some stage 'recruiting further down the pool of applicants will become cost ineffective as a way of increasing the numbers enrolled at tertiary level in SET with any reasonable chance of success' (p. 162).

In Chapter 2 it was pointed out (see 2.5) that entry-level assessment has the potential to smooth the transition between secondary and tertiary study, by identifying levels of competence and appropriate support.

Although the results in the Mathematics Access Module improved from 2002 to 2003, it is not possible to conclude that Video 2 or the associated reading activities provided in the first assignment in 2003 brought about the improvement. It was also evident that even though the video may have had an impact, the number of students admitted to the examination was low. Only 40% of those who registered finally wrote the examination: many students could not even obtain sufficient credits to qualify for the examination. (Table 8.4 in Chapter 8 reflects the increasing attrition rate: from 43% in 2001, to 54% in 2002 and 65% in 2003.) Each intervention introduced to open up access to mathematics seemed to uncover additional problems, and it became clear that without the prerequisite skills success would continue to elude students. Diagnostic assessment seemed to be the only way of finding out whether these skills were in place.

Entry-level testing is a familiar practice, internationally and in South Africa. In Chapter 2 some aspects of international practice (see 2.5.1) and local practice (see 2.5.2) are noted. In 2003 pre-registration assessment was already an accepted practice at many tertiary institutions, and possibility that assessment mechanisms used elsewhere could be applied at UNISA was considered. It was important that such assessment should not be regarded as a gate-keeping device, but as a means of obtaining information to identify and advise potentially 'at risk' students. It was also important (until the validity of the process could be established, and widespread acceptance gained) that students should not be expected to pay for any form of diagnostic assessment, so that they would not be excluded on financial grounds. The ultimate purpose of such assessment was that students would be accurately placed in three different categories, namely students likely to need more support than the university was in a position to provide ('high-risk' students); students who would benefit from support in specific areas, such as reading remediation, in order to be better equipped to study mathematics (‘medium-risk’ students); and the ‘low-risk’ students, who could be expected to have the potential to cope on their own, using the study guides, tutorial letters and such tutorial classes as were available, if required.

It was evident that there would be large numbers of potential Access Module students in the medium-risk category who would need considerable support. Such students needed to be identified as early as possible. Various support options could then provide sufficient scaffolding to ensure greater chances of success for these students.

In Chapter 2 (see 2.5.1) it was noted that any form of testing, whether for selection or for placement, is likely to create certain tensions, especially in the context of open and distance learning. However, the potential benefits appeared to outweigh such considerations.

### 9.3 Acceptance of diagnostic assessment for the Mathematics Access Module

Testing at entry level is used to provide students with advice, offer early placement, provide guidance as to choice of courses, and in cases where numbers are limited, guide the institution in student selection. A significant difference between UNISA and other institutions is that for UNISA, economies of scale favour large enrolments, whereas at face-to-face teaching institutions enrolment is limited by physical facilities. The necessity for testing within the Mathematics Access Module arose from the need to offer better advice to students, and to inform both the university and potential students about the types of support that are required and/or available.

In June 2003 a proposal[1] was put to the UNISA Senate to introduce compulsory diagnostic assessment for all potential Mathematics Access Module students. The Senate approved, seeing the proposed assessment as a means of helping students to identify current strengths and weaknesses (in terms of academic and quantitative literacy) and of providing better guidance regarding appropriate directions of study, or determining appropriate support options.

It was agreed that assessment should be based on internationally accepted standards, adapted if necessary to suit specific UNISA requirements. Issues of reliability and validity were to be considered in the selection of assessment tools, as well as the need for practical and cost-effective measures. The diagnostic assessment was not approved as a *prerequisite* to study, but as a

[1] Formulated by Carol Bohlmann.

compulsory co-registration requirement. It was thus possible to offer advice to students, but not to apply the assessment results rigorously in terms of acceptance into or exclusion from the Mathematics Access Module.

## 9.4 Investigating various options

Before decisions were taken regarding the selection of an appropriate diagnostic assessment tool, various processes being used elsewhere in South Africa were investigated. The two most significant ones are discussed below.

### 9.4.1 Testing at the University of Cape Town

***The Alternative Admissions Research Project (AARP)***

The AARP tests (see also 2.5.4 in Chapter 2) consist of a 'Placement Test in English for Educational Purposes' (PTEEP), a 'Mathematics Comprehension Test' (MCom) and a 'Mathematics Achievement Test' (MAch). The English and the Mathematics Comprehension Tests assess ability to learn rather than knowledge previously learnt; the Mathematics Achievement Test relies heavily on subject content knowledge (Cliff, Visser, Hanslo & Yeld, 2003). In these tests, results are analysed and adjustments made according to a set formula to smooth out differences in school background.

***A trial of the AARP tests at UNISA***

Before study material was written for the Mathematics Access Module the possibility was considered of allowing potential students to write the AARP tests and thereby qualify for admission to the Bridging Module (Precalculus A) and one other first-year module (Precalculus B) without having to go via the Mathematics Access examination that existed at that time (1997). This examination (see Chapter 4) tested Grade 12 content knowledge at SG level. In order to make a decision, a trial run of AARP testing was carried out at UNISA, with students registered for the Mathematics Access examination. An agreement with UCT was reached, whereby UCT supplied UNISA with AARP scripts (at a fee of R65,00 per student), marked the scripts and analysed the results, making adjustments for school background.

In August 1997 a group of 47 students who were registered for the Mathematics Access examination and were living within reasonable travelling distance of UNISA were invited to write the three AARP tests during the course of one day, from 09:30 to 16:00, with an hour's break for lunch. Of the students who were contacted, 16 agreed to write the three tests. Their results, together with the mark they received in the October Mathematics Access examination, are shown in Table 9.1.

**Table 9.1:**

**Results of UNISA students in AARP tests and the Mathematics Access examination**

| **Script number** | **PTEEP (%)** | **PTEEP (decile)*** | **MAch (%)** | **MAch (decile)** | **MCom (%)** | **MCom (decile)** | **Math Access exam (%)[2]** |
|---|---|---|---|---|---|---|---|
| 9…8 | 9,8 | 10 | 10 | 10 | 2,6 | 10 | (13) |
| 9…9 | 22,2 | 7 | 26, | 7 | 5,3 | 10 | 27 |
| 9…0 | 19,0 | 8 | 23,3 | 8 | 5,3 | 10 | 33 |
| 9…1 | 32,0 | 5 | 20 | 9 | 14,6 | 8 | 23 |
| 9…3 | 23,5 | 7 | 40 | 4 | 32 | 4 | 27 |
| 9…4 | 22,9 | 7 | 20 | 9 | 12 | 8 | 17 |
| 9…5 | 4,6 | 10 | 16,6 | 9 | 10,6 | 8 | 10 |
| 9…6 | 7,8 | 10 | 30 | 6 | 5,3 | 10 | 20 |
| 9…7 | 15,0 | 9 | 53,3 | 3 | 12 | 8 | 37 |
| 9…8 | 17,6 | 8 | 30 | 6 | 13,3 | 8 | 23 |
| 9…9 | 28,8 | 6 | 20 | 9 | 14,6 | 10 | 23 |
| 9…0 | 24,8 | 7 | 20 | 9 | 5,3 | 10 | (13) |
| 9…1 | 23,5 | 7 | 26,6 | 7 | 6,6 | 5 | 27 |
| 9…2 | 27,5 | 6 | 33,3 | 6 | 24 | 8 | (23) |
| 9…3 | 29,4 | 6 | 30 | 6 | 14,6 | 8 | 30 |
| 9…4 | 38,6 | 4 | 20 | 9 | 12 | 8 | 27 |
| Mean | 21,7 | | 26,2 | | 11,8 | | 23,3 |

*'Decile' refers to the stratification of students into percentage groups, where the 10th decile refers to the bottom 10% when compared to all other students tested.

The mean marks (percentages) for English (PTEEP), Mathematics Achievement (MAch) and Mathematics Comprehension (MCom) were respectively 21,7%, 26,2% and 11,8%. For all three

[2] The percentages in brackets in this column refer to the June examination mark. These students were repeat students who had registered again for the second semester, but for some reason did not write the examination in October.

of the tests, 14 of the 16 the students were in the lower five deciles, i.e. in the bottom half of students when compared to UCT applicants. In their interpretation of the results, UCT used as an example the scores of one particular student (script number 9…8, first row) who had attended an ex-Department of Education and Training (DET)[3] school. With regard to his PTEEP mark (9,8%) and PTEEP decile (10), the UCT report stated that

> The PTEEP decile referred to here is that established from the scores of the ex-DET UCT applicants who wrote this test. In other words, if Mr M… had written the test as an applicant to UCT, he would have been in the bottom 10% when compared to UCT's other applicants who matriculated at a school formerly classified as DET. When we make recommendations to Deans, they are on the following basis: Science Faculty: we would recommend applicants who were placed in the top four deciles for ex-DET – this would be for placement onto our Science Foundation Programme. This appears to be a very weak group, not only in terms of their Maths performance (UCT report, private communication, 1998).

Only one student in the English test, and two students in the Mathematics Achievement Test (one of whom was similarly placed in the Mathematics Comprehension Test) reached the top four deciles.

These 16 students all scored better in the MAch than in the MCom (the group means were 26,2% for achievement and 11,8% for comprehension). However, the items used in the Mathematics Achievement Test were generally considered to be at too advanced a level to be used for potential UNISA students (if it was assumed that UNISA would continue to admit students who had not necessarily studied mathematics beyond Grade 9).

The AARP trial suggested that the tests had predictive value for success for potential UNISA students. The majority of these students would correctly have been classified as weak students who were unlikely to be successful in their studies, certainly at UCT, and as the examination results indicated, at UNISA as well, even at the level of the Mathematics Access Module.

For a number of reasons UNISA decided to continue with the development of study material, and not pursue the idea of an entry or selection test for potential Access Module students. However, this information shows that the kinds of questions included in the AARP tests (particularly the

[3] Before 1994 schooling was separate, for separate racial groups. The majority of African school children attended schools administered by the Department of Education and Training. See Chapter 2.

focus on English for academic purposes, and on mathematics comprehension) appeared to have predictive value for Mathematics Access Module students at UNISA. For this reason, when diagnostic testing was being considered in 2003, it seemed logical to consider whether UNISA could not simply make use of the AARP tests already in existence. The AARP tests were recognised as theoretically relevant to UNISA students, but for practical reasons it was decided that other options should be considered. As a distance-teaching university UNISA needs to be able to offer the diagnostic assessment at any of its examination centres, anywhere in the world. The logistics involved in coordinating delivery of scripts by UCT to UNISA's examination centres, and collecting and returning the scripts to UCT for processing, would be enormous. Furthermore, AARP tests are not electronically available, and although electronic testing might now be foreign to many students, UNISA needs to move towards a position of being able to provide computerised testing in the future.

### 9.4.2 Testing for placement at the University of Port Elizabeth

***Background to placement testing at UPE***

UPE was established in 1969 as a bilingual (English and Afrikaans) university. In 1997 its student population comprised a diverse mix of students from different population groups (39% black, 16% Indian and coloured, and 45% white; by 1999 the percentage of black students had increased to 52% (Foxcroft, 1999)). UPE draws most of its students from the Eastern Cape, a relatively economically impoverished region (GCIS, 2003). At entry level many of its students have similar characteristics to UNISA students (such as variable levels of schooling, low socio-economic status, diversity of language and culture, registration several years after leaving school), and it seemed that much could be learnt from the UPE experience.

As a face-to-face institution, several of the UPE aims at the time of beginning to develop placement testing were not equally applicable to the UNISA situation (for example, the idea that test results could be used to provide lecturers with information regarding students' entry-level needs and competencies so that individual adaptations to a student's study programme could be made). However, several aspects of the initial vision of the team involved in developing placement testing at UPE (for *all* prospective first-year students) were relevant to UNISA, for example the idea that such testing would be a responsible means of ensuring a reasonable chance of success for the students it accepted, and could lead to placement of students in appropriate programmes according to individual needs and competencies. The UPE view was that a test battery would need to take into account cultural, language and educational background; recognise potential, rather than acquired knowledge and skills; meet standard psychometric criteria (i.e. provide credible measurements of student cognitive competence); be researched on an ongoing basis to establish predictive validity, reliability and suitability, and be checked for bias (Foxcroft, 1999). With time the competencies that appeared to be most closely related to academic success at UPE were academic literacy (i.e. familiarity with and facility in the use of academic discourse), English language proficiency, and numeracy, as well as certain non-cognitive factors, such as motivation and post-school experience (Foxcroft, 1999).

Since 1999 UPE has assessed entry-level competencies using a combination of matric results, an algebra and an arithmetic test, and a reading comprehension test. The algebra, arithmetic and reading tests are incorporated in a battery of tests known as ACCUPLACER CPTs (Computerised Placement Tests) obtained from the Educational Training Service (ETS) in the USA. These tests were originally developed by the ETS, but then became the property of the College Board of the USA. ACCUPLACER assesses English language proficiency, English reading, basic algebra and basic arithmetic. The ACCUPLACER algebra, arithmetic and reading tests are computerised adaptive tests. In the computer version of ACCUPLACER, item response theory is used to generate a unique test for each test taker. ACCUPLACER tests can be obtained via the Internet from the College Board's agent and non-adaptive printed versions (called the COMPANION) can be purchased from them in booklet form for students unfamiliar with computers, or in situations where computerised testing is not possible.

***Success of ACCUPLACER at UPE***

At UPE the test results were used to establish a risk profile for each student, on the basis of which academic performance could be predicted, developmental needs could be established, and early remedial action could be taken (Koch & Foxcroft, 2002).

From 2000, all prospective students have been tested, regardless of the faculty for which admission was sought, or of school-leaving results. Admissions and placement assessment programme (APAP) reports (APAP Report: Profiles of the 2002 and 2003 first-year intakes) (Foxcroft, Watson & Seymour, 2002; Foxcroft, Seymour & McSorley, 2003), which deal with the academic performance of students assessed at UPE, were made available for UNISA to consider. From the UPE results it appears that admissions and placement assessment has the potential to provide institutions with critical information that can be used to address issues such as the need for increased access, and the improvement of throughput rates, and at the same time adopt a learner-centred, developmental approach to facilitating learning (Foxcroft, Watson, Davies & Beneke, 2002). Over the years, admissions and placement testing appears to have been successful in reducing the number of high-risk students admitted to various degree programmes. Ongoing research has made it possible for UPE to move from local testing only to testing at remote sites; to investigate the effects on students and on test results of computerised and written versions of tests; to research the practicalities of using ACCUPLACER (either as the tests are, or in an adapted form), and to investigate a number of other issues.

## 9.5 Practical aspects of introducing diagnostic assessment at UNISA

### 9.5.1 Investigating the assessment process

In order to determine whether diagnostic assessment would be acceptable to various stakeholders within the UNISA community, a number of discussions were held over a period of about two months prior to Senate approval being granted, and subsequently.

Discussions were held with members of staff of the Institute of Continuing Education (ICE), which is responsible for the coordination of all access modules; with members of staff in the Departments of Practical English, Linguistics, Mathematics, Computer Science, Statistics, and Student Support; with staff members in the Bureau for Learning Development (BLD), the Bureau for Student Counselling and Career Development (BSCCD), the Examination Department, the Marketing Departments and the Department of Undergraduate Student Affairs.

Students are more likely to benefit from an assessment if they are able to have the assessment results in time to take appropriate action, if necessary. At UNISA registration commences in December, but can continue until the end of February or early in March. This long registration period means that students may not yet have registered by the time the assessment is written. It was thus decided to set two assessment dates, one in January and the other mid year, even though later assessment would not be ideal. Two registration codes[4] were created for the assessment, to accommodate the two different test dates.

The Marketing Departments (in the Science Faculty, as well as the for the university as a whole) were asked to disseminate all relevant information as widely and as frequently as possible, so that all potential students could be made aware of the new procedure. Information was also included in the UNISA Calendar and the Access Brochure, which provide general information pertaining

[4] These codes were MDA010-F (January) and MDA011-G (June).

to the university as a whole, and information of particular relevance to students taking access modules, respectively.

When students registered for the assessment they received an information letter (Tutorial Letter 101[5] for MDA010-F/MDA011-G) explaining all relevant aspects of the assessment. From the UPE experience it was clear that such an explanatory letter was necessary to deal with the many questions students were likely to have when registering for the assessment, as well as to provide examples of the types of questions that could be asked, which would show students that it was not possible to study for the assessment.

### 9.5.2 Assessment format

For a number of reasons computerised testing was not viable for 2004. However, opportunities for computerised testing need to be further investigated, as this would probably be the most efficient and effective means of carrying out diagnostic assessment in a distance-learning environment. Ideally diagnostic assessment needs to be readily available, at any UNISA centre, at any time, provided staff at Learning Centres and Regional Centres can make the necessary arrangements, i.e. set up appropriate test times, arrange for adequate computer availability and supervision, and assist students with practical details.

### 9.5.3 Costs of the assessment

Students were not asked to pay a fee for this assessment. Clearly the introduction of diagnostic assessment would involve additional cost for the university. Having made the decision that these costs would not be passed on to the students the financial implications of the process were not taken into account.

## 9.6 Choice of assessment tools

In the USA

> … college admissions counselors know that the student who shows high capability in *both* the mathematics and verbal sections of the SAT is more likely to succeed in math than the student who has a severely skewed score, strong only in quantitative skills (Tobias, 1987, p. 5).

Chapters 5 and 6 highlighted the extent to which poor reading skills may be a barrier to mathematical performance. It was decided that even though access to *mathematics* was the issue,

[5] See Appendix H.

diagnostic assessment needed to focus on both reading and quantitative skills, as precursors to the development of appropriate mathematical skills. (See also 2.5.3 in Chapter 2.)

### 9.6.1 Selection of the ACCUPLACER Reading Comprehension

The advantage of using existing tests, even though norms and standards might differ, is that these tests have already been screened for validity and reliability. As has been discussed, the AARP tests were not appropriate for logistical reasons. The option of computerised delivery and marking of the ACCUPLACER tests was seen as a significant long-term potential advantage for UNISA even though it was clear that computerised testing was initially not viable.

The costs involved in purchasing the ACCUPLACER tests (either the printed tests, or the 'right' to be tested, if the online version of the ACCUPLACER tests was selected) need to be taken into account. Such costs would be USA dollar-related, and could escalate to unaffordable levels. However, such costs (and hence the opportunity to provide students with appropriate advice and preempt failure) need to be weighed against the costs (to the students and to the university) of registering students who fail or drop out. For 2004 the printed ACCUPLACER tests were purchased from funds remaining in the Mathematics Access Module account[6].

The decision was made to focus on reading rather than language proficiency. Reading effectively is fundamental to constructing meaning from printed mathematics texts. The ACCUPLACER Reading Comprehension (ARC) was selected, as it was seen to be more important than the ACCUPLACER English Language Proficiency test for potential Mathematics Access Module students. The ARC is accompanied by a marking schedule, and raw scores are converted to scores which distinguish between three categories of students (see 9.8.1). UPE had used the ACCUPLACER tests in their American format. An empirical item bias investigation was being undertaken to see which items needed to be adapted. Before a bias analysis could be undertaken it was necessary that large numbers of students were tested on the items. For UNISA, the ARC conversion table and suggested categories were used as guidelines.

An analysis of the ARC items identified several different aspects of reading that were tested (see Table 9.2). Four questions were regarded as potentially problematic with regard to ambiguity of words and phrasing. For example, one question asked how the following two sentences are related:

---

[6] Established at the time of a previous Dean, Prof G McGillivray.

*Automobile gas mileage will be the foremost consideration of American buyers from now on.*

*Now that all cars sold in America have achieved a certain level of fuel efficiency, buyers will choose on the basis of performance.*

The following four options were given, with (a) the correct answer.

(a) The second contradicts the first.

(b) The second expands on the first.

(c) The second is a consequence of the first.

(d) The second is a restatement of the first.

This question and three others were also seen as examples of particularly culture-specific items, where the terminology clearly reflects North American culture. In spite of possible problems arising from ambiguities and culture-specific items, members of staff in the Departments of Linguistics and Practical English felt that the American bias did not detract from the ability of the items to assess students' ability to construct meaning from text. In an academic context students will in any case often be required to make sense of unfamiliar passages, possibly in unfamiliar contexts. As these tests were used internationally, and were statistically assessed in terms of validity and reliability, it was decided to leave the items in their original format for the first round of testing, and delete items if the test results suggested that this was necessary.

### 9.6.2 The development of a test for quantitative reasoning

The level of content knowledge required for potential Mathematics Access Module students is considerably lower than the standard of the items in the mathematical sections of many other placement or diagnostic tests, such as the AARP Mathematics Achievement Test, and the ACCUPLACER Basic Arithmetic and Basic Algebra tests. An important aspect of the AARP Mathematics Comprehension Test was that it did not draw on previously taught mathematics, but primarily tested reasoning potential, using limited mathematical knowledge that students could realistically be expected to have at the end of secondary school. Since the AARP tests were not a practical option for UNISA, it was necessary to design a quantitative reasoning test for UNISA's purposes that would include similar types of items. This gave rise to the Basic Arithmetic Test (BAT). An important aspect of the AARP tests is that they are dynamic tests, i.e. they include an element of teaching, in order to measure students' potential to learn and apply what has been taught. Time constraints limited the extent to which this could be done in the BAT, and only two items were included in which students could learn from some form of teaching in the question. (See Table 9.3.)

Students who have not understood the rudimentary concepts taught up to the end of Grade 9 mathematics have little chance of understanding the Mathematics Access Module content, and it was necessary to determine whether sufficient understanding of Grade 9 mathematics had been reached, as well as to assess whether potential students could use their knowledge in simple problem solving exercises. The design of the test[7] took these factors into account, and drew upon the types of problems experienced by Mathematics Access Module students. Questions in the assessment were based on conceptual issues that appear to cause problems for students, and hamper further understanding.

In general, questions varied in difficulty from simple computational tasks to activities requiring insight, reasoning skills, and the ability to interpret simple sentences mathematically. For example, in the reading skill section included in the first assignment in 2003, students had been asked to note the time at which they began reading, the time they completed their reading, and the total reading time (the difference between starting time and finishing time). A number of students *added* the two times together. This gave rise to the following question in the diagnostic assessment:

> *Maria starts to read a magazine article at 8:30. She finishes the article at 10:00. How long does it take her to read the article?*

Several other questions were included to probe students' understanding of time (with respect to 'when' as well as with respect to 'how long'), such as

> *Joseph paints a cupboard in* 2½ *hours. Sydney paints twice as fast as Joseph. How long does it take Sydney to paint the cupboard?*

Experience with Access Module students suggested that many students would assume that the correct answer to this question was five hours.

An important aspect in comprehending mathematical concepts is an ability to recognise patterns. This is emphasised in discussions regarding the nature of mathematics, for example 'Maths is pattern' (Freeman, 1994, p. 11) or 'Mathematics is a science, "the science of patterns"' (MSEB, 1993, p. 22). Three pattern recognition questions were included, for example a question in which

[7] The test was designed by the module leader for the Mathematics Access Module, Carol Bohlmann, with input from other mathematics lecturers.

students were required to recognise the pattern involved in computing the number of dots in a block.

In order to construct meaning, students must be able to make sense out of phrases containing comparisons such as ‘twice as fast as’, ‘at least’, ‘greater than’, and so forth. Although understanding phrases such as these may appear to relate to language proficiency rather than quantitative reasoning, they are inseparable from the mathematical aspects to which they relate, as can be seen from the painting example above. Several questions testing this aspect were thus included.

The most important aspect of designing a diagnostic test is to ensure that the test items do actually reveal strengths and do not simply ‘skate over the surface’ (Mason, 2002, p. 108). Mason points out that it is easy to overlook ways in which students might get the right answer for the wrong reason. With all diagnostic testing, a student’s reasoning is as important as the answer, which suggests that multiple-choice items may not be the best tool. For practical reasons it may be the only one available. With MCQs it is never possible to know whether students obtained the correct answer by guessing, although random placing of the correct option minimises this risk.

Since reliability and validity of the items could not be ensured in the same way as for the ARC, the BAT was scrutinised by various members of staff who had had considerable experience with Mathematics Access Module students, and members of staff from the university’s Bureau for Learning Development, who had had experience in test design, bias in testing, and so forth.

### 9.6.3 Administering the assessment

Students were required to register for the Mathematics Access Module and other approved modules *and simultaneously* for the diagnostic assessment. Mid-year assessment allows less time for any remedial action, and for this reason students were strongly advised to register early, and wait for the results before registering for any modules in the Science Faculty. However, this was not possible for students who registered after the first assessment had been written.

The ARC test booklets (Part 1 of the assessment) were couriered to UNISA by ACCUPLACER; the BAT paper (Part 2 of the assessment) was produced by UNISA. It was decided that, since the ARC contained 35 questions, the BAT would also consist of 35 questions. All were multiple-choice questions, each with four options. All questions were answered on the same mark-reading sheet. The reading items were numbered 1 to 35 and the arithmetic items 36 to 70. Students received both parts at the same time, could tackle questions in any order, and were allocated three hours for the entire assessment. Space was available on the question papers for rough work.

Each student received an assessment package, containing the ARC booklet, the BAT paper, a mark-reading sheet, and a list of instructions. Invigilators were also issued with a similar list of instructions, emphasising the importance of returning all question papers together with the mark-reading sheets, primarily for security reasons, as the same tests were to be used in the second phase of the assessment. For research purposes it was essential that the students undertook the same assessment.

## 9.7 Description of the assessment components

### 9.7.1 ACCUPLACER Reading Comprehension (ARC)

An analysis of the ACCUPLACER test items identified different aspects of reading. See Table 9.2.

**Table 9.2:**

**Aspects of reading tested in the ACCUPLACER Reading Comprehension (ARC)**

| Aspect tested* | Number of occurrences** |
|---|---|
| Causal relations | 2 |
| Contrastive relations | 4 |
| Recognition of sequence | 1 |
| Interpretation of implied or inferred information | 9 |
| Comprehension/interpretation of factual information; detailed/ general. | 5 |
| Identification of main/subsidiary ideas | 12 |
| Academic vocabulary | Many (in 33 questions) |
| Number sense | 1 |
| Recognition of contradiction, inconsistency | 1 |
| Substantiation | 2 |

* The UNISA reading project had identified anaphoric resolution as an important skill (see Chapter 5), but this was not directly assessed in the ARC. Since accurate anaphoric resolution is central to comprehension it was perhaps not essential that this aspect be directly assessed; however an investigation into the results may lead to the inclusion of specific anaphoric items at a later stage.
** There were 35 questions, but the total is greater than 35, as several questions tested more than one aspect.

### 9.7.2 Basic Arithmetic Test (BAT)

The questions were grouped into categories, as shown in Table 9.3.

**Table 9.3:**

**Aspects of quantitative reasoning tested in the Basic Arithmetic Test (BAT)**

| **Aspect tested** | **Number of occurrences**** |
|---|---|
| Simple arithmetic operations (whole numbers) | 19 |
| Simple arithmetic operations (fractions/decimals) | 12 |
| Pattern recognition | 3 |
| Number sense | 5 |
| Conversion of units | 3 |
| Academic vocabulary (e.g. satisfy) | 3 |
| Technical vocabulary (e.g. zero, fraction, ratio) | 4 |
| Comparison (e.g. bigger than, earlier, more than half, double) | 10 |
| Time (how long) | 3 |
| Time (when) | 2 |
| ‘Translation’ from words to mathematical statements | 17 |
| Recognition of insufficient or redundant information | 4 |
| Learning from explanation | 2 |
| Spatial awareness | 4 |
| Insight | 14 |

**There were 35 questions but the total is greater than 35, as several questions tested more than one aspect.

These categories are fairly arbitrary and subjective, and often overlap. Consider for example Question 58 (placed in two categories: ‘simple arithmetic’ and ‘spatial awareness’).

> *The sketch shows a volcano, which rises 1 800 m above sea level. The volcanic crater is 300 m deep. Water has collected in the crater, and the water level has risen to 10 m at the deepest point. A tourist drops his map into the crater, and it floats on the surface of the water. How high above sea level is the map?*

A sketch of the volcano is given, and the students can select from four options: 1 490 m, 1 510 m, 2 090 m and 2 110 m. It would be difficult to determine whether a student obtained the correct answer by simply adding and subtracting according to the use of the words ‘rise’ and ‘depth’, or whether the student actually had any three-dimensional sense of the situation described.

‘Number sense’ was an issue in the painting example mentioned earlier, where students were asked

> *Joseph paints a cupboard in* 2½ *hours. Sydney paints it in 2 hours. How long does it take them to paint the cupboard if they work together?*

Many students doubled the time, and did not realise that two people working together should take less time. However, meta-cognition is an additional factor to consider here, in that students often do not question the validity of the answers they obtain.

The first ten questions were presented in numerical format only, for example

35 ½ x 10 =

followed by the four options. All other questions required some facility with reading and an understanding of the words used to convey the question, for example a question in which students were required to arrange numbers in ascending order (although the question was simplified by using the words ‘from smallest to biggest’, in case ‘ascending’ was unfamiliar). Where ‘translation’ is mentioned as one of the aspects tested, additional reading skills were required. Translation refers to questions where specific phrases needed to be converted to some mathematical operation. Another aspect tested, namely insight, was often related to comparison or spatial awareness, and was thus difficult to separate from these aspects. For example, students needed to recognise that leaving a place *earlier* implied traveling *longer*; they also needed to have a sense of the fact that *doubling* the length and breadth of a photograph would lead to an enlargement with an area *more than double* that of the original. Insight was also related to the recognition of insufficient information, although these were listed as two separate categories.

### 9.8 Grading criteria

Criteria needed to be set for the purpose of categorising students who were considered likely to be successful in the Mathematics Access Module with no additional assistance (Category 53)[8], those who were considered likely to be successful provided they had support (Category 52), and those who were unlikely to be successful without assistance beyond that available at the university

[8] Details of the categories (51, 52 and 53) and the implications of this classification were explained to students in Tutorial Letter 101 for MDA010-F/011-G, which students received when they registered for the assessment. See Appendix H.

(Category 51). UPE had developed a profile across all the scores in the battery (including school performance) and used regression formulae and classification functions to classify potential students as being high, medium or low risk. These formulae were developed on the basis of previous research. At UNISA the process of setting criteria for classification was based on the ARC guidelines, the research results emanating from the reading intervention, and consensus from staff in the Mathematics Department who had been involved with Mathematics Access Module or in the planning of the diagnostic assessment project). These categories may be refined as the process continues.

### 9.8.1 Criteria for the ARC

For the Reading Comprehension the ACCUPLACER Coordinator's Guide[9] (ACCUPLACER, 2002) provided a table to convert the raw score out of 35 to a score out of 120. A minimum raw score of 0 converted to a final score of 20, and a maximum raw score of 35 to a final score of 120. The ARC recommendations on the Reading Comprehension, based on their research, provided three categories, viz. the weakest with a score of about[10] 51 out of 120[11]; a moderate category with a score of about 80 out of 120; a good score of about 103 or higher.

These scores reflect various levels of reading skill. Students who obtained in the region of 51 (out of 120) could comprehend short passages characterised by uncomplicated ideas, and straightforward presentation, in which subject matter generally reflects everyday experience; distinguish between main and subsidiary ideas; recognise the tone of a passage when fine distinctions are not required; and recognise relationships between sentences.

Students who obtained a score of about 80 could comprehend short passages characterised by moderately uncomplicated ideas and organisation; answer questions that require synthesis of information, including gauging point of view and intended audience; recognise organising principles used in the text; and identify contradictory or contrasting statements.

Students with scores of 103 or more could comprehend more complex passages (in terms of ideas) dealing with academic subject matter, often in a theoretical framework; extract points that

[9] See Appendix I for the ACCUPLACER conversion table.

[10] The ACCUPLACER guide uses this terminology, referring to scores of 'about' 51, etc. since the conversions do not make it possible to score exactly 51, 80 or 103. See Appendix I.

[11] Clearly there would have be some students who scored less than 51; they would presumably constitute a small minority who are clearly not functionally literate, and who would probably not have been expected to take the test.

are merely implied; follow moderately complex arguments or speculations; recognise tone; and analyse the logic employed by the author in making an argument.

From the conversion table it is clear that some form of weighting of questions takes place. All scores from 0 to 4 out of 35 are converted to 20 points, i.e. any student with between no and four correct answers would obtain a score of 20 points. The increment between the numbers of correct answers increases gradually from 1 point to 4 points. This can be seen in the ACCUPLACER conversion table, given in Appendix I. It would seem that students are 'rewarded' for obtaining a greater number of correct answers.

In the ACCUPLACER guide the upper boundary of the lowest category was 80. A total correct score of 80 on a scale from 20 to 120 is roughly 67%. Phases I and II of the reading project described earlier suggested that a score of 60% appeared to be a threshold below which reading skills were too weak to support effective study. In the diagnostic assessment it was thus decided to consider students with less than 60% on the Reading Comprehension as being at risk with respect to reading. For the MDA the upper boundary of the lowest category was pitched at 70 out of 120 (which is actually 58%). A score of 70 out of 120 is equivalent to a raw score of 22 or 23. If a student obtained 22 or 23 correct answers out of the 35 questions the converted scores would be 67 and 71, respectively, out of 120 (corresponding to approximately 56% and 59%, respectively). Pitching the MDA boundary at 70 seemed to be justifiable: the ARC and the reading tests used in Phases I and II (described in Chapter 6) are different but many aspects of the two tests are similar and there appears to be some similarity between the threshold described in Phases I and II and the limits set in the ARC.

### 9.8.2 Criteria for the BAT

The BAT questions were weighted according to the perceived degree of difficulty.
Questions with weights of 1 or 1,5 were classified as 'easy', and were the simplest, procedural types of questions (such as addition or multiplication). Questions with weights of 2 or 2,5 were classified as 'moderately easy' and 'moderately difficult', and required, in addition, some insight and interpretation of language. Questions with a weight of 3 were classified as 'difficult', and required, in addition, greater insight and interpretation of language.

The maximum total correct score was thus 69 and not 35. The weighting was intended to enable students who demonstrated greater insight to score better than those who had merely memorised procedures without understanding. Table 9.4 shows the questions (Q) and respective weights (W).

**Table 9.4:**

**Weights (W) of Basic Arithmetic Test questions (Q)**

| Q | W | Q | W | Q | W | Q | W | Q | W | Q | W | Q | W |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 36 | 1 | 41 | 1,5 | 46 | 2 | 51 | 3 | 56 | 2 | 61 | 3 | 66 | 1 |
| 37 | 1 | 42 | 1,5 | 47 | 3 | 52 | 2,5 | 57 | 3 | 62 | 2 | 67 | 2 |
| 38 | 1 | 43 | 1,5 | 48 | 3 | 53 | 2,5 | 58 | 2,5 | 63 | 2 | 68 | 1,5 |
| 39 | 1 | 44 | 1,5 | 49 | 2,5 | 54 | 2,5 | 59 | 1,5 | 64 | 2 | 69 | 2 |
| 40 | 1 | 45 | 1 | 50 | 2 | 55 | 2 | 60 | 3 | 65 | 2 | 70 | 2,5 |

Table 9.5 shows the distribution of questions with different weights, and the contribution of the different categories to the total.

**Table 9.5:**

**Respective proportions of weighted questions in the Basic Arithmetic Test**

| Category | Weight | Percentage of total | Number of items |
|---|---|---|---|
| Easy | 1 | 10 | 7 |
| Easy | 1,5 | 13 | 6 |
| Moderately easy | 2 | 30 | 10 |
| Moderately difficult | 2,5 | 22 | 6 |
| Difficult | 3 | 25 | 6 |

It can be seen that students could score up to 53% by correctly answering easy to moderately easy items (with weights of 1, 1,5 or 2); they could score up to 75% by correctly answering ‘easy’, ‘moderately easy’ and ‘moderately difficult’ items (with weights up to 2,5). Students who scored over 75% also needed to answer the ‘difficult’ items (with weights of 3) correctly.

In the BAT ten items (with weights accounting for 17% of the total) were purely computational in nature and required no reading skills. It seemed reasonable to expect most students to answer these questions correctly. There were thus 25 items (accounting for 83% of the total) that depended to some extent on reading, as well as on other aspects such as comparison and pattern recognition. If one considers the 60% ‘threshold’ identified with respect to reading skill, it seemed reasonable to set 17% + (60% of 83%), i.e. 17% + 50%, or 67%, as the lowest boundary for this component of the assessment. This was equivalent to a raw score of 46 out of 69. Students with scores of less than 46 were thus considered to be at risk with respect to numerical and quantitative reasoning. At the other end of the spectrum, the top category in the Reading Comprehension comprised scores of 103 (approximately 86%) or higher. To achieve such scores in the BAT, students needed to score 59 or more out of 69. With no other empirical evidence available this was selected as the cut-off point for high achievement in the BAT.

### 9.8.3 Categorisation

Table 9.6 was used to place student scores into the three different categories.

**Table 9.6:**

**Diagnostic assessment categories**

| **Category** | **Reading Comprehension** | | **Basic Arithmetic** |
|---|---|---|---|
| 51 | score < 70 | **OR** | score < 46 |
| 53 | score $\geq$ 103 | **AND** | score $\geq$ 59 |
| 52 | All other scores | | |

The following examples illustrate this table.

| | | | |
|---|---|---|---|
| ARC score 68 | and | BAT score 55: | Category 51 |
| ARC score 74 | and | BAT score 41: | Category 51 |
| ARC score 75 | and | BAT score 61: | Category 52 |
| ARC score 103 | and | BAT score 60: | Category 53 |

## 9.9 Results of the assessment

### 9.9.1 Practical aspects

Several administrative problems such as late registrations and students not being given examination centre information in time resulted in only 326 students taking part in the assessment in January 2004. The rest wrote later (in June 2004, when 834 students were assessed).

Once the assessment had taken place the mark-reading sheets and question papers were sent to UNISA from the different examination centres. The Assignment Department then captured all the marks from the mark-reading sheets. These marks were sent to the Department of Computer

Services, where a programme had been written to convert the raw scores into the three categories according to the pre-determined formula described in Table 9.6. Apart from the conversion to categories, the only other analysis undertaken involved item-test correlations, which were computed taking individual items in relation to the test as a whole[12]. See Tables 9.7 and 9.8.

Once the assessment had taken place, a number of issues became apparent. In one of the BAT questions (Question 65) every student obtained the correct answer, which seemed unrealistic, even though it was a ‘moderately easy’ question, with a weight of 2 (this did not occur in any other question, even ‘easy’ questions with a weight of 1). There was no obvious reason why this should have happened, as there was no mistake in the options given for the question. There may have been some programming error, which needs to be investigated. This question was thus discarded before further analysis of the results took place.

There were several students who obtained zero for the ARC, and a number who obtained zero for the BAT, which was unexpected. Subsequent examination of the mark-reading sheets showed that some students had not answered any of the questions in one of these sections. Invigilators were on hand to answer questions during the assessment, so it is unlikely that students did not know how to fill in the mark-reading sheets. It is possible that students may have filled in the answers on the question papers; however the blank sections of the mark-reading sheets were discovered too late to try to trace the question papers and see whether questions had been answered on the question papers instead of on the mark-reading sheets.

### 9.9.2 ARC results

The item-test correlations for the ARC are shown in Table 9.7.

**Table 9.7:**

**Reading Comprehension results (*n* = 326)**

| **Question** | **No. correct answers** | **Item-test correlation *t*** (to two decimal places) | ***$t^2$*** (to two decimal places*) |
|---|---|---|---|
| 1 | 232 | 0,46 | 0,21 |
| 2 | 220 | 0,57 | 0,32 |
| 3 | 159 | 0,89 | 0,80 |

[12] The correlations were first done by the Department of Computer Services, at the same time as the conversion of raw scores to categories.

| | | | |
|---|---|---|---|
| 4 | 269 | 0,37 | 0,13 |
| 5 | 184 | 0,59 | 0,35 |
| 6 | 160 | 0,67 | 0,45 |
| 7 | 114 | 0,75 | 0,56 |
| 8 | 68 | 0,35 | 0,13 |
| 9 | 187 | 0,61 | 0,37 |
| 10 | 245 | 0,39 | 0,15 |
| 11 | 196 | 0,66 | 0,43 |
| 12 | 172 | 0,77 | 0,59 |
| 13 | 239 | 0,50 | 0,25 |
| 14 | 181 | 0,55 | 0,30 |
| 15 | 150 | 0,64 | 0,41 |
| 16 | 160 | 0,91 | 0,83 |
| 17 | 187 | 0,54 | 0,29 |
| 18 | 247 | 0,47 | 0,22 |
| 19 | 58 | 0,46 | 0,22 |
| 20 | 132 | 0,22 | 0,05 |
| 21 | 134 | 0,42 | 0,18 |
| 22 | 238 | 0,57 | 0,33 |
| 23 | 215 | 0,69 | 0,48 |
| 24 | 224 | 0,57 | 0,32 |
| 25 | 180 | 0,71 | 0,51 |
| 26 | 145 | 0,49 | 0,24 |
| 27 | 200 | 0,75 | 0,56 |
| 28 | 202 | 0,62 | 0,38 |
| 29 | 150 | 0,68 | 0,47 |
| 30 | 222 | 0,60 | 0,36 |
| 31 | 218 | 0,69 | 0,48 |
| 32 | 187 | 0,70 | 0,49 |
| 33 | 109 | 0,81 | 0,66 |
| 34 | 81 | 0,48 | 0,23 |
| 35 | 158 | 0,25 | 0,06 |

* Calculations of $t^2$ were based on the original *t* values, before rounding (given correct to 5 decimal places).

Calculations of $t^2$ show, for example, that 83% of the variation in Question 16 can be associated with the students' total scores. Correlations were moderate on many of the items that had been regarded as potentially problematic (Questions 1, 5, 8, 9, 12, 15, 16, 19 and 26). Only three of these questions (1, 8 and 19) had low (i.e. below 0,5) item-test correlations, with, for example, only 13% of the variation in Question 8 associated with the students' final scores.

### 9.9.3 BAT results

The item-test correlations were weaker in the BAT than in the ARC. This is understandable, since several of the BAT items tested aspects unrelated to each other: for example, there is little expectation that being able to accurately compute the addition of two three-digit whole numbers (Question 36) is conceptually related to determining the number of dots in the 50th block in a set of nested blocks where a picture is given showing the number of dots in the first three blocks (Question 61). The results are shown in Table 9.8.

**Table 9.8:**

**Basic Arithmetic Test results (*n* = 326)**

| **Question** | **No. correct answers** | **Item-test correlation *t*** (to two decimal places) | $t^2$ (to two decimal places*) |
|---|---|---|---|
| 36 | 294 | 0,14 | 0,02 |
| 37 | 304 | 0,11 | 0,01 |
| 38 | 255 | 0,25 | 0,06 |
| 39 | 251 | 0,26 | 0,07 |
| 40 | 130 | 0,40 | 0,16 |
| 41 | 169 | 0,64 | 0,41 |
| 42 | 138 | 0,54 | 0,29 |
| 43 | 194 | 0,63 | 0,40 |
| 44 | 203 | 0,25 | 0,06 |
| 45 | 242 | 0,37 | 0,14 |
| 46 | 252 | 0,39 | 0,15 |
| 47 | 209 | 0,30 | 0,09 |
| 48 | 200 | 0,56 | 0,32 |
| 49 | 210 | 0,50 | 0,25 |
| 50 | 311 | 0,10 | 0,01 |
| 51 | 219 | 0,47 | 0,22 |
| 52 | 190 | 0,74 | 0,55 |
| 53 | 240 | 0,28 | 0,08 |

| | | | |
|---|---|---|---|
| 54 | 130 | 0,51 | 0,22 |
| 55 | 258 | 0,16 | 0,26 |
| 56 | 166 | 0,63 | 0,40 |
| 57 | 183 | 0,66 | 0,44 |
| 58 | 155 | 0,73 | 0,53 |
| 59 | 277 | 0,26 | 0,07 |
| 60 | 163 | 0,61 | 0,37 |
| 61 | 88 | 0,24 | 0,06 |
| 62 | 288 | 0,24 | 0,06 |
| 63 | 123 | 0,51 | 0,26 |
| 64 | 232 | 0,46 | 0,21 |
| 65 | 326 | - 0,00 | - |
| 66 | 228 | 0,20 | 0,04 |
| 67 | 91 | 0,22 | 0,05 |
| 68 | 243 | 0,32 | 0,10 |
| 69 | 136 | 0,47 | 0,22 |
| 70 | 125 | 0,30 | 0,09 |

*Calculations of $t^2$ were based on the original *t* values (given correct to 5 decimal places), before rounding.

Low (below 0,30) item-test correlations were obtained across all question categories: for several of the ‘easy’ questions, for example questions 36, 37, 38, 39, 66 (which all carried a weight of 1) and 44 (weight = 1,5); for several of the ‘moderately easy’ and ‘moderately difficult’ questions, for example questions 50 (with a weight of 2, which had the lowest item-test correlation), 53 (weight = 2,5), 55, 62 and 67 (weight = 2), as well as for some of the ‘difficult’ questions, such as questions 47 and 61 (weight = 3).

Calculations of $t^2$ show, for example, that 55% of the variation in Question 52, and 10% of the variation in Question 50, respectively, can be associated with the students’ final scores.

### 9.9.4 Consolidation of results

Once the results were consolidated, the allocation of the January students to the three different categories yielded the following:

| | | |
|---|---|---|
| Students in Category 53 | | 10 |
| Students in Category 52 | | 93 |
| Students in Category 51 | | 223 |
| | Total | 326 |

The ratio 223/326 shows that approximately 68% of the students who participated in the assessment were perceived to be at risk (Category 51). Interestingly, in the combined October 2003 and January 2004 Mathematics Access Module examinations (consolidating the year-end and supplementary examination results) 76,34% of the students failed, suggesting that they may have been ‘high-risk’ students, had they been similarly assessed. It is thus perhaps not surprising that a large proportion of the 2004 students performed poorly in the diagnostic assessment in January.

Preliminary analysis of the June results revealed that in spite of a few months exposure to the Mathematics Access Module study material (and in many cases also to the study material for the English Access Module), allocation of students to the different categories yielded similar results:

| | | |
|---|---|---|
| Students in Category 53 | | 35 |
| Students in Category 52 | | 176 |
| Students in Category 51 | | 623 |
| | Total | 834 |

Roughly 75% of the students who were assessed in June could thus have been considered to be ‘high-risk’ students (Category 51).

Further analysis of the January and June scores (i.e. the sum of the scores of all the students who participated in the assessment) reveals an interesting pattern: the unweighted ARC mean (26,64%) is less than the unweighted BAT mean (32,36%); however the gap between the weighted ARC mean (48,47%) and the weighted BAT mean (62,28%) is greater.

A comparison of the means of the total scores of the two groups (raw and weighted) indicates that in spite of more than three months exposure to an academic environment there was little difference between two groups. The results are shown in Table 9.9.

**Table 9. 9:**

**Comparisons of mean scores of the January and June groups**

| Group | Number of students | Raw score mean (total) | Raw score mean (%) | Weighted score mean (total) | Weighted score mean (%) |
|---|---|---|---|---|---|
| January | 326 | 41,31 (70) | 59,01 | 103,07 (189) | 54,53 |
| June | 834 | 41,30 (70) | 59 | 100,37 (189) | 53,11 |

Further analysis showed that although these differences are small they are statistically significant. The results suggest that many students have been unable to benefit from exposure to tertiary-level study. A comparison of the diagnostic assessment results and examination results may yield further information.

### 9.10 Discussion of results

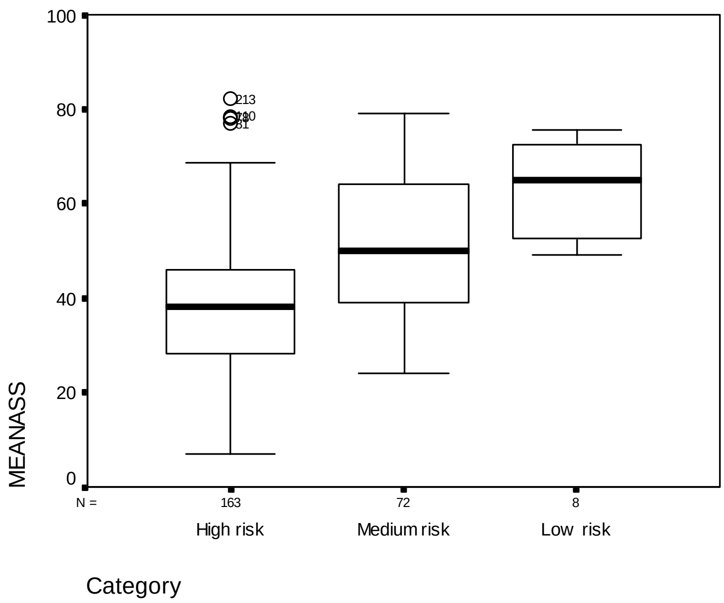
The consolidation and conversion of the results provided risk categories. Item-test correlations were also carried out. Later in the year the results were analysed in a little more detail. Initially the item-test correlations were done with individual items against the test as a whole. These correlations were repeated within the subject domains. Although the correlations changed slightly, the trend remained the same, showing higher correlations with respect to the reading questions. As it was pointed out earlier, a number of students answered only Part 1 or Part 2, and not both. Missing answers were processed as incorrect answers, i.e. 0 was awarded to questions in which the answer was incorrect, and in cases where no answer had been given. In cases where the entire section was omitted this will have affected the overall results. Ideally, the results need to be calculated again distinguishing between these two cases.

It would also have been helpful to know why students left out sections, if indeed they did so. Subsequent to the initial analysis of the results a request was put to the Bureau for Learning Development (BLD) to conduct focus group interviews with selected students from the three score categories, and from specific groups, such as those who only answered one part of the assessment. It was also suggested that postal questionnaires could be used to gather student perception of the diagnostic testing process and the availability of the necessary support. Although there was initial interest in these suggestions, the BLD did not pursue the matter.

The performance of the January students in the different categories was also compared to their assignment marks. Figure 9.1 represents the results by means of a box plot, with a few outliers shown separately. The vertical axis shows the mean assignment marks of the students in each category. This is also a problematic measure, in that many students submit only enough assignments to ensure that they have examination admission, and a student with high scores on two assignments (and hence more than 100 credits) is not necessarily comparable with a student who obtains moderate scores on four assignments. At UNISA in many subjects there is often a poor correlation between assignment and examination marks. The small number of students in the low risk category also diminishes the value of this information, but the trend is nevertheless interesting.

**Figure 9.1**

**Comparison of mean assignment scores across risk categories (January)**




Anecdotal evidence emerging during the year has also shown that students[13] in Category 51 on the whole least understood the purpose of diagnostic assessment, and were experiencing many problems. For example, several Category 51 students assumed that 51% denoted ‘passing’ the diagnostic assessment, and hence admission to the examination, although the use of 51 as a categorisation symbol had been clearly explained.

[13] Note that these findings relate to the students who made appointments to see the module leader, sent e-mails or phoned. It cannot be assumed that they reflect all students in Category 51.

Once examination admission had been finalised it was evident that many of the students who had not obtained examination admission, in spite of submitting all seven assignments, were in Category 51 (hardly surprising, since this category included the majority of the students). In contrast, several students in Category 52 or 53 contacted the module leader during the course of the year to find out how they had performed in the two different parts of the assessment. Several (in Category 53) were concerned that they had scored less well than they had expected.

The fact that the gap between weighted ARC/BAT scores is wider than the gap between the unweighted scores (for the whole group) suggests that the weightings of the different questions should be revisited. The scoring of the BAT items possibly needs to be further investigated as well. Studies using the Newman Error Analysis[14] have shown that the error profiles for students with identical scores can be completely different. There were certain BAT items that tested more than one aspect of reasoning, arithmetic and/or interpretation, and a simple numerical provision of 1 or 0 for right or wrong, respectively, even when the questions are weighted, makes no allowance for individual differences in problem-solving approach (Clements & Ellerton, 1996).

With regard to the potentially problematic ARC questions, the item-test correlation for Question 8 was the lowest, followed by Questions 1 and 19. Performing well in these questions was thus unlikely to predict the overall test result in any way. It may be useful to recalculate the category classification without these three questions, even though student performance was poor in only two of them (Questions 8 and 19).

Analysis of the January BAT results shows that of the items with moderate to good correlation (higher than 0,4) there were twelve which less than two thirds of the students answered correctly. It may be valuable to revisit these questions with specific groups of students. The BAT showed markedly weaker item-test correlations than did the ARC, suggesting that the individual reading items were more strongly associated with overall performance than many of the individual BAT items. The reading scores may thus have played a greater role in determining the categories into which students were placed. This could possibly have been expected, since a significant number of the BAT items also required a certain level of reading proficiency.

[14] A model based on the work of an Australian language educator in the late 1970’s but later seen to be relevant for mathematics as well. The model provided a systematic procedure for analysing errors made by students responding to written mathematical activities. Analysis of data based on Newman procedures have highlighted the influence of language on mathematics learning, the importance of ‘mathematising’ (see Chapter 4), and the need to avoid equating remedial mathematics with revision of algorithms (Clements & Ellerton, 1996).

It is clear that there is a need for further research into the results and the individual test items to refine the tests and work towards establishing their predictive validity. It is also important that the quantitative research be supported by additional qualitative research.

### 9.11 Diagnostic assessment as part of an identification-action cycle

If testing is used to diagnose problems in order to help students deal with them, the provision of support is critical, otherwise the assessment could become a means of excluding students, rather than placing them in appropriate courses. When students registered for the assessment they received written information explaining the purpose of the assessment and the meaning of the different categories. They were encouraged to see the assessment as a means of helping them identify weaknesses, so that appropriate action could be taken before it was too late.

#### 9.11.1 Category 51

Students in Category 51 were strongly advised not to study the Mathematics Access Module in 2004; however, if they chose to ignore the advice, UNISA was perhaps in a morally sounder position than before. It is of course possible that students in this category would have been sufficiently challenged to take advantage of whatever remedial programmes were available, and put in additional effort to be successful, despite predictions to the contrary.

One of the support departments, the Institute for Continuing Education (ICE), seemed best placed to assist Category 51 students. The ICE had agreed that if any of the ‘high-risk’ students contacted them (as was suggested to students in the information letter accompanying registration for the assessment) they would offer help regarding alternative directions of study, or measures that could be taken to upgrade academic skills *before* attempting university study. The ICE had initially considered appointing an additional member of staff to deal with the increased workload; however, this was never necessary, as very few students contacted them. Although it was theoretically possible that students in Category 51 would seriously reconsider their decision to study the Access Module in 2004, the fact that Undergraduate Student Affairs required *simultaneous* registration for the diagnostic assessment and the Mathematics Access Module resulted in very few students taking this advice seriously. Cancellations in the early part of the year did not appear to be more numerous than at similar times in previous years.

#### 9.11.2 Category 52

There was a perception (not substantiated by the January or June results) that the majority of students would be placed in Category 52, and that such students would need substantial support, in a number of areas. In an early meeting discussing diagnostic assessment, the National Tutorial Support Coordinator was in favour of the process, as it appeared that the information obtained would help the Department of Student Support (DSS) provide more effective tutorial support. The Bureau for Student Counselling and Career Development (BSCCD) staff were hesitant as to their ability to deal with the problems on an ongoing, wide-scale basis, but willing to assist where possible. The Povey Centre, an independent centre providing (for a fee in most cases) instruction in English language proficiency, reading and writing skills, felt able to commit to a degree of reading instruction at some of the Learning Centres, for students in need of such assistance. They expected to be able to provide this at no additional charge (other than the basic Learning Centre registration fee required from all students who registered for tutorial classes at the Centres).

Various support options, among them the following, were thus suggested for students placed in Category 52. Firstly, an explanation of the category and the implications of this categorisation for Access Module study, which would be provided by a designated mathematics lecturer (telephonically, or in person on the Main Campus). Secondly, additional assistance with respect to reading mathematics: this support could take the form of facilitated viewing of the 'Read to Learn' video and video workbook activities (involving tutors at the Learning Centres). This could assist students with their reading in a specifically mathematical context. Additional tutorial support with respect to Access Module content could also be provided at the Learning Centres. From Chapter 4 it can be seen that such classes were provided from 1999, but it was envisaged that more specific attention could be given to students in Category 52 (as well as to those in Category 51 who chose to continue with their mathematics studies).

Thirdly, the English Access Module staff had previously developed additional activities for those of their students who needed additional help. Mathematics students studying both the mathematics and the English modules could also do these activities; their progress could be monitored and they could be assisted to transfer what they were learning in their English module to their mathematics module. This could take place at the Learning Centres, involving tutors in both the English and Mathematics Access Modules. Students could also be assisted to upgrade their reading skills at the Learning Centres, with instruction provided by staff appointed by the

Povey Centre. Clearly these activities were dependent on additional tutor training and the resources required to provide it.

Suitably qualified staff in the Bureau for Student Counselling and Career Development (BSCCD), at UNISA Regional Centres, could give appropriate guidance to students uncertain about their choice of subjects. Advice with regard to other educational options could also be available from ICE staff (probably only telephonically, or in person on the Main Campus).

#### 9.11.3 Category 53

Students in Category 53 were expected to cope with their studies, and no suggestions were made for additional support for these students. They were of course welcome to use whatever support was provided.

### 9.12 Conclusion

The costs, benefits, advantages and disadvantages of instituting diagnostic assessment for Mathematics Access Module students at UNISA need to be thoroughly investigated. Information regarding the whole implementation process, and the data that become available, need to be well utilised. In the interests of true access it is crucial that students are given appropriate career guidance. The implementation of diagnostic assessment at UNISA could serve as one of the tools in this process.

### 9.13 Summary

In this chapter we have reviewed the final intervention described in this thesis, and the associated action research cycle. This cycle was the culmination of the sequences of cycles in the two parallel sets of action research cycles discussed in Chapters 5, 6 and 7.

In this chapter the rationale for the introduction for diagnostic assessment was given. Once the proposal to introduce diagnostic assessment for Mathematics Access Module students had been accepted, several options were investigated, the two most significant being the AARP tests and the ACCUPLACER tests. Practical aspects of introducing the MDA at UNISA were discussed. We then considered the reasons for selecting the ACCUPLACER Reading Comprehension, and for developing a UNISA-specific test for quantitative literacy. Administration of the test was discussed, and the test components were described. We then considered the grading criteria used in the ARC and BAT, and the consolidation of scores into risk categories. MDA results for

January and June were given separately, as well as the consolidated results for both groups together. Analysis of results was of a limited nature, providing opportunity for further research once examination results can be considered as well.

The envisaged support options are also discussed, and it is emphasized that the MDA should form part of an identification-action cycle, eliminating risk and increasing the potential for success.