A Distribution Model for the Assessment of Database Systems Knowledge and Skills among Second-Year University Students

Linda Meiring

Dissertation submitted in fulfilment of the requirements for the degree

MAGISTER TECHNOLOGIAE:
INFORMATION TECHNOLOGY

in the

SCHOOL OF INFORMATION AND COMMUNICATION TECHNOLOGY
Faculty of Engineering, Information and Communication Technology

at the

Central University of Technology, Free State
January 2009

Supervisor: Dr H.J. Vermaak
Co-Supervisor: Mr P.H. Potgieter

DECLARATION OF INDEPENDENT WORK

I, LINDA MEIRING, identity number ▮▮▮▮▮▮▮ and student number 20349335, do hereby declare that this research project submitted to the Central University of Technology, Free State for the degree MAGISTER TECHNOLOGIAE: INFORMATION TECHNOLOGY is my own independent work; that it complies with the Code of Academic Integrity, as well as other relevant policies, procedures, rules and regulations of the Central University of Technology, Free State; and that it has not been submitted before to any institution by myself or any other person in fulfilment (or partial fulfilment) of the requirements or the attainment of any qualification.

_____   _____

L. Meiring   DATE

ACKNOWLEDGEMENTS

I would like to express my gratitude to:

- Jesus Christ, my Saviour and reason for living

- Herman Vermaak, my supervisor, for his input

- Pieter Potgieter, my co-supervisor, for reading through every chapter and willingly giving his input

- Anton, my husband, for his encouragement

- Prof. Antoni Szubarga, for his assistance with the statistics

- Erica Wessels, for the linguistic revision of the dissertation

- All others who provided assistance in any form

SUMMARY

Assessment is a method used by primary, secondary and tertiary institutions to determine whether the child or student mastered the necessary skills needed to be successful in a specific subject or course. These assessments are conducted by an assessor. The assessor needs to know what types of test items to use in order to successfully test a person.

Research was done to determine the quality of database test items. The quality of test items in a test will determine the accuracy by which the assessor can determine if a person successfully mastered a subject or course.

Test items can be presented to students in different ways. Through Computer Based Assessment (CBA) students are tested online. The five different types of CBA test-models were discussed of which a combined model were used during the investigation.

The assessor will draw up test items to test the database knowledge of the student. A distribution model was identified to demonstrate the current distribution of test items on each of Bloom's Taxonomy Levels

Test item keywords were identified and the curriculum modules and assessment outcomes were linked to the subject areas of Microsoft Access. Test items were identified as either single- or multi-level test items and were again linked to a specific object, curriculum outcome, cognitive level and difficulty level. It seemed that a single- and/or multi-level test item normally determines the cognitive level of a test item.

A number of existing online test-generating software packages, which allows the assessor to draw up test items, was investigated. Criteria were identified whereby these packages could be evaluated. It became clear that there is a

need for test-generating software packages that can create single- and multi-level test items. It must be possible to link these test items to a database object, difficulty level and cognitive level, as well as curriculum outcomes.

Students were assessed on single- and multi-level test items. The analysis of the results revealed a relation between the single- or multi-level test item, the object, and the difficulty level of the test item. A revised distribution model was suggested for lecturers teaching databases to use as a guideline when drawing up test items.

A software product known as TestGen was developed. This program will help the assessor to create single- and multi-level test items and to link these items to a difficulty level, cognitive level and curriculum outcome.

Further research is suggested in view of creating a model that will help assessors to draw up test items for second- and third-year database students. It is also suggested that TestGen be revised in order to automatically generate tests. Test items must be randomised based on the difficulty level, cognitive level and curriculum outcomes selected.

OPSOMMING

Eksaminering is 'n metode wat gebruik word deur primêre, sekondêre en tersiêre instansie om te bepaal of die kind of student die nodige vaardighede bemeester het om suksesvol te wees in 'n spesifieke vak of kursus. Die eksaminering word deur 'n eksaminator uitgevoer. Die eksaminator moet weet watter tipe toetsitems om te gebruik om 'n persoon suksesvol te kan toets.

Navorsing is gedoen om die kwaliteit van databasisitems te bepaal. Die kwaliteit van toetsitems in 'n toets sal die akkuraatheid bepaal waardeur die eksaminator kan bepaal of 'n persoon die vak- of kursusinhoud suksesvol bemeester het.

Toetsitems kan op verskillende manier aan die student voorgehou word. Deur middel van Rekenaargebaseerde Assessering word studente aanlyn getoets. Die vyf verskillende tipes Rekenaargebaseerde Assessering is bespreek waarna 'n gekombineerde model gebruik is gedurende die navorsing.

Die eksaminator sal toetsitems opstel om die student se databasiskennis te toets. 'n Verspreidingsmodel is geïdentifiseer om die huidige verspreiding van toetsitems op elk van Bloom se Taksonomie Vlakke te demonstreer.

Toetsitemsleutelwoorde is identifiseer en die kurrikulummodules en assessering uitkomste is gekoppel aan die vakareas van Microsoft Access. Toetsitems is identifiseer as enkel- of multivlak-toetsitems en is weer gekoppel aan 'n spesifieke objek, kurrikulum-uitkoms, kognitiewe vlak en moeilikheidsgraad. Dit wou voorkom asof enkel- en/of multivlak-toetsitems gewoonlik bepaal wat die kognitiewe vlak van 'n toetsitem sal wees.

'n Aantal bestaande sagteware produkte wat toetsitems genereer en die eksaminator toelaat om toetsitems op te stel, is nagevors. Kriteria is geïdentifiseer waarmee pakkette geëvalueer kan word. Nadat 'n paar pakkette

geëvalueer is, het dit duidelik geword dat daar 'n behoefte bestaan aan sagteware pakkette wat enkel- en multivlak-toetsitems kan skep. Daar is ook 'n behoefte geïdentifiseer om 'n toetsitem te kan koppel aan 'n databasisobjek, 'n moeilikheidsvlak, kognitiewe vlak en kurrikulum-uitkomste.

Studente is geëksamineer aan die hand van enkel- en multivlak-toetsitems. Die analise van die resultate het getoon dat daar 'n verband bestaan tussen enkel- of multivlak-toetsitems, die objek en moeilikheidsgraad van die toetsitem. 'n Hersiene model is voorgestel wat dosente in die databasisvak kan gebruik as riglyn wanneer toetsitems opgestel word.

'n Sagteware produk met die naam TestGen is ontwikkel. Die program sal die eksaminator help om enkel- en multivlak-toetsitems op te stel en te koppel aan 'n moeilikheidsgraad, kognitiewe vlak en kurrikulum-uitkoms.

Verdere navorsing is voorgestel om 'n model daar te stel wat eksaminatore sal help om toetsitems op te stel vir tweede- en derdejaar databasis studente. Daar is ook voorgestel dat TestGen hersien moet word sodat die program toetse outomaties sal kan genereer. Toetsitems moet ewekansig geselekteer word op grond van die moeilikheidsgraad, kognitiewe vlak en kurrikulum-uitkomste wat geselekteer is.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

α:	Significance
CBA:	Computer-Based Assessment
CBT:	Computer-Based Test
CurVerb:	Curriculum Verb
CUT:	Central University of Technology, Free State
ECDL:	European Computer Driving Licence
ETM:	Easy Test Maker
IQS:	Interactive Question Server
IR:	Item Response
IT:	Information Technology
INL10DB:	Information Systems (first-year students)
INL20DB:	Information Systems (second-year students)
INL30EB:	Information Systems (third-year students)
L1:	Level 1 (Easy)
L2:	Level 2 (Medium)
L3:	Level 3 (Difficult)
LOFT:	Linear-on-the-Fly Testing
M2:	Multi-Level 2
M3:	Multi-Level 3
OLE:	Object Linking and Embedding
PDF:	Portable Document Format
P&P:	Paper and Pencil
QMA:	Question Management Area
QuestionTbl:	Question Table
S1:	Single-Level
SD:	Standard Deviation
STEM:	Science, Technology, Engineering and Mathematics
SUT:	Sentrale Universiteit vir Tegnologie, Vrystaat

TDS: Test-Delivery System

TMA: Test-Management Area

TMS: Test-Management System

WYSIWIG: What You See Is What You Get

XML: eXtensible Markup Language

CHAPTER 1 : INTRODUCTION

1.1   GENERAL

Assessment is a daily occurrence. Employees are assessed at their workplaces, and students are assessed at the institutions at which they are studying. But why is assessment necessary?

Genesee and Upshur [1] identified different aspects of classroom-based evaluation (Figure 1.1). They stated that there must be a clear purpose as to why an assessor wants to track the progress of students, and this purpose must be clearly communicated to the students in question. If the purpose is to determine whether an instruction method is effective or if students can be promoted to the next level, there must be a means whereby this can be assessed. Students at the Central University of Technology, Free State (CUT) are assessed to determine whether they have mastered the important skills needed to be successful in a specific course. Students are also evaluated to determine whether they can be promoted to a higher level.

In order to assess a particular student, certain information is needed. Such information can be collected by different means, such as the assessment of group work, as well as observation, individual projects, tests and so on.

Once the necessary information has been gathered, the data must be interpreted. The results of the data can be used by the assessor to determine whether a particular student has reached an acceptable level of knowledge to allow him/her to proceed to a higher level. The results can also be used to determine whether the student's progress is on track and to guide the student in improving his/her performance. The assessor can also determine whether a different instruction method should be used or whether there should be greater focus on specific areas of the work not successfully mastered by the student.
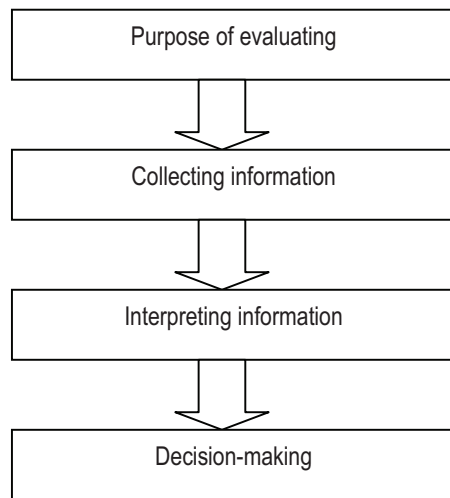
Figure 1.1: Aspects of classroom evaluation

A well-known assessment method used at the CUT is the paper-and-pencil (P&P) based testing method. McDonald's [2] description of a P&P-based test is that it is normally a hand-out page with questions that the student should attempt to answer.

There are advantages to using the traditional P&P-based method. Kwosek [3] reported that all the instructors that were questioned for purposes of his research were of the opinion that the P&P method could better evaluate students' comprehension than computer-based assessment (CBA) tests. CBA, also known as computer-based testing (CBT) or e-exam, is an electronic version of a handwritten exam, with an electronic device like a computer being used to assess students [2]. This assessment method can also give feedback to the student as well as the assessor. In 2002 McDonald [2] predicted that CBA would play a significant role in assessment and feedback in the future.

Kwosek [3] identified certain advantages to using computer-based assessment instead of P&P testing, namely:
- The instructor's work is minimised;
- Time spent on grading students is reduced;

- o Human error is eliminated and accuracy improved; and
- o The incidence of cheating by students is reduced.

According to Zakrzewski and Bull [4] a lack of resources such as hardware, software, technical expertise, security and reliability can have a negative effect on CBA. Bocij and Greasley [5] stated that a lack of a backup system for student responses, when students are writing online tests can result in data loss in the event of a network failure.

If the above is taken into consideration, then personal circumstances and resources will determine whether CBA or P&P-based testing will be used. Any assessor deciding to use CBA must research the existing software available that can be of assistance in this regard. Software can be used to create test items, save test items to a test bank, draw up question papers, assess papers, analyse test results, and save user results [5].

Thomas [6] described a test item as a stimulus towards which a student will respond and which will be evaluated or scored. When CBA is used to test students, a bank of test items is needed. An item bank is a carefully calibrated set of test items that "develop, define and quantify a common theme and thus provide an operational definition of a variable" [7].

Test items must successfully test whether a student is ready to be promoted to the next level of study. Test-item distribution models are used to help determine the quality of test papers [8]. In this study, the focus is on a distribution model for INL20DB students doing Microsoft Access. Available software that allows the assessor to create test items, store the test items in a test bank and draw up test papers is evaluated.

1.2   INTRODUCTION

One of the mainstream subjects for second-year Information Technology (IT) students at the CUT is Information Systems II (INL20DB). The CUT curriculum for second-year IT students stipulates the content that must be mastered by a student in a specific subject. The curriculum for second-year IT students in the subject INL20DB specifies that the student must be competent in the use of Microsoft Access 2002 as database software.

A database is usually a large collection of data that is organised in such a way that it is easy to conduct a search to retrieve the right information [9]. In order to assess a student's knowledge of Microsoft Access, the assessor can give the student a printed test paper instructing the student to apply his/her database knowledge on computer, and the assessor will then assess the student's database knowledge manually. Another method would be to assess the student in a virtual environment where, instead of receiving a paper, the student receives online test items in a Microsoft Access environment and applies his/her knowledge to an online database. Computer software would then be used to assess the database of each student.

CBA can be used as formative or summative assessment. An assessor normally starts with formative assessment in order to determine the knowledge level of a student in a specific course [10]. At the end of a programme or module, summative assessment is used to check the level of learning and whether the student has indeed achieved the specific outcomes of that programme or module. If the results of the formative assessment are compared with the results of the summative assessment, this may produce evidence of whether or not the student has in fact mastered the curriculum content [11]. The focus in this study is on summative assessment.

According to researchers at the Oshkosh College of Business [12] test items in a test bank are characterised by a number of parameters, one of which may be that test items must match the outcomes in a curriculum and that these items must then build on one another in the same hierarchical way that the curriculum does. In order to ensure that test-bank items test what they are suppose to test, another parameter, namely item analysis, must be applied. Item analysis is a tool that ensures test effectiveness by studying a student's response to each test item. An item can be measured according to item difficulty, item discrimination, and effectiveness of alternatives. Item difficulty will determine how difficult the student found an item to be. Item discrimination measures the effectiveness of an item in differentiating between students who have mastered a skill versus those who have not. Item analysis also measures the effectiveness of alternatives, where a test item has three possible answers, of which a student must select one, leaving two effective distracters.

Item analysis also implies content validity. Content validity in item-bank development exists where there is a correlation between the curriculum items and the thinking skills within that specific course [12].

Test items can be categorised to show the student's mastery of different thinking skills. Bloom [13] created a learning taxonomy in 1956 to categorise assessment test items. According to Bocij and Greasley [5], in 1999 the majority of assessment packages at that stage were incapable of testing or developing cognitive skills at the higher analysis, synthesis and assessment levels of Bloom's taxonomy.

1.3 RESEARCH GOALS AND OBJECTIVES

Existing software packages that allow an assessor to create test items will be evaluated. Existing models [14], [15], [16] will be used to evaluate these software packages, with the focus on:

- o Quality
    - o Software functionality: Is the software working correctly and effectively?
    - o Software efficiency: How quickly does it generate tests and results? What is the extent of the hardware resources used?
    - o Software portability: Is it easy to install?
    - o Software usability: Is the software user-friendly?

- o Layout and Functions
    - o Functions: Are functions like Find/Replace, Spellchecker and Help available?
    - o Is paper compilation possible?
    - o Can test items and answers be edited?
    - o Can a certain number of test items be randomly selected from the databank?
    - o Exporting and printing of the paper and memorandum: Can the test items be exported to another program like Microsoft Word if one wants to make changes or add formatting like numbering? Is it possible to print the paper directly from the assessment package?
    - o CBA: Is it possible to translate the test items into a computer-based assessment test?

- o Databank Items
    - o Is it possible to group items according to subject matter or curriculum outcomes?
    - o Are there different types of test items, like True/False? How many types are available?
    - o Is it possible to create single- and multi-level test items?
    - o Is it possible to group items according to Bloom's taxonomy?

In this study, the term "multi-level test items" refers to test items that build on one another. A level-one test item is an item that a student can answer/do without any preceding task(s). A level-two test item builds on a level-one test item. The level-one test item must be answered prior to answering the level-two test item, and so the second part of the question can only be answered once the first part of the question has been answered. Levels three, four, etc. work on the same basis.

The focus of this study is on the creation of a test-item distribution model for INL20DB students, along with software that will allow the assessor to draw up single- and multi-level test items.

It is easy to select test items randomly from a databank when all the items are at level one. It becomes more difficult in the case of multi-level test items, because the test items are linked to one another. Randomisation minimises cheating by students, which results in a more accurate assessment of students. Randomisation is therefore an important characteristic of computer-based assessment [10] and must be taken into account.

1.4 PROBLEM STATEMENT

Research has been conducted on the development of software packages to enable an assessor to draw up test items. Packages like Articulate Quizmaker 2.0 [17], ExamWeb [18], WebQuiz [19], CourseWebs [20], Respondus [21] and many more allow the assessor to create test items and save these in a databank. Articulate Quizmaker 2.0, for instance, allows the assessor to choose between 21 different types of test items.

For purposes of this research, different packages are evaluated to determine the following:

- o The possibility of drawing up multi-level test items;
- o The different types of test items that can be drawn up, e.g. Yes/No and True/False test items;
- o The ability to randomise test items;
- o The ability to link a test item to the programme outcomes stipulated in the curriculum;
- o The ability to link a test item to one of Bloom's taxonomy skills; and
- o The ability to link a test item to a difficulty level.

When programme outcomes [22] are tested, test items are evaluated against the following:

- o The conditions under which a student must be able to do something, e.g. use Microsoft Access;
- o Tasks that the student must be able to complete, e.g. save a database;
- o The minimum level of performance, e.g. be able to successfully save a database;
- o The ability to link a test item to one of Bloom's taxonomy levels.

The question arises as to whether the second-year IT students' knowledge of Microsoft Access is properly tested if they are evaluated on single-level test items alone. Second-year IT students at the CUT are not assessed online, because available online assessment tests are expensive. An example of an online CBA test is the European Computer Driving Licence (ECDL) [23], where students subscribe to the service and write online tests on, for instance, Microsoft Access, using the ECDL databank test items. These test items are mainly level-one test items.

For purposes of this research, second-year IT students were evaluated on level-one and multi-level test items. An attempt was made to determine the distribution of single-level versus multi-level test items for second-year students, based on the following:

- The difficulty level of the test item;
- Bloom's taxonomy levels; and
- The object (table, query, form, macro).

1.5   RELATED WORK

There are commercial packages available that can be used by assessors to draw up tests. Commercial products such as Articulate Quizmaker 2.0 from TechLearn [17], ExamWeb from Assessment Technologies [18], WebQuiz from SmartLite Software [19], CourseWebs [20] and many more are available. There are also some free packages available, such as Respondus [21].

The test items developed by the assessor are saved in an item bank. Software will then help the assessor to generate tests from these item banks. The test can be printed for a P&P test, or students can write these tests online in a virtual environment.

Computer-based assessment (CBA), where students register to write online tests, offers enormous scope in testing and evaluating students [10]. The European Computer Driving Licence (ECDL) Foundation offers online CBA tests with existing item banks [23]. When the student completes one of these online tests, the system will mark the student's responses. This is expensive if one considers that the institution has to pay for every test written by the student. This can result in each student writing only a limited number of tests. Besides being expensive, another disadvantage is that the assessor does not have access to these item banks and is forced to follow the ECDL curriculum.

The advantage of an online test like the ECDL is that the student writes the test in a virtual environment. Thus, if a student writes a Microsoft Access test, he/she applies his/her knowledge on the computer within the virtual environment of Microsoft Access.

1.6   HYPOTHESIS

A test-item distribution model, along with test-generating software, will assist the assessor in drawing up single- and multi-level test items based on the curriculum outcomes, subject area, difficulty level, and Bloom's taxonomy levels.

1.7   HYPOTHETICAL RESOLUTIONS

A test-item distribution model will be developed that will help the assessor to draw up test items to test whether the student has successfully mastered the curriculum outcomes. A test-generating software package will be developed that allows an assessor to develop single-level and multi-level test items to be saved in a databank. These assessment items will be grouped according to the curriculum outcome, object, difficulty level, and Bloom's knowledge level. It will be possible to generate a paper-based test with the test-generating software.

1.8   METHODOLOGY

The study consists of 5 phases:

Phase 1: Identifying the different test models (software-measuring instruments)

Phase 2: Determining the curriculum outcomes for INL20DB students at the CUT, as well as their levels in terms of Bloom's taxonomy.

Phase 3: Evaluating available software packages that can draw up test papers.

Phase 4: Data analysis of single- and multi-level test items given to second-year IT students.

Phase 5: Developing a software package that will allow the assessor to draw up multi-level test items linked to a difficulty level, one of Bloom's taxonomy levels, an object and a course outcome.

1.9   SUMMARY

The purpose of this research is to develop a test-item distribution model for INL20DB students, along with software that will allow the assessor to draw up single- and multi-level test items. Existing test-generating software packages will be investigated to determine whether test items can be linked to a difficulty level, Bloom's taxonomy levels, an object, and the curriculum outcomes for that specific subject.

In Chapter 2, existing test models are investigated along with software-measuring instruments.

1.10 REFERENCES

[1]   Genesee, F. & Upshur, J.A. (1996).   *Classroom-based evaluation in second-language education*. Cambridge: Cambridge University Press.

[2]   McDonald, A.S. (2002). *The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments.* Computers and Education, 39(3): 299-312.

[3]   Kwosek, C.T. (2001). *Assessment of computer-based evaluation as used by vocational technical instructors*. Menomonie, WI: University of Wisconsin-Stout.

[4]   Zakrzewski, S. & Bull, J. (1998). *The mass implementation and evaluation of computer-based assessments.* Assessment and Evaluation in Higher Education, 23(2): 141-152.

[5]   Bocij, P. & Greasley, A. (1999). *Can computer-based testing achieve quality and efficiency in assessment?* International Journal of Educational Technology, 1(1): 1-18.

[6]   Thomas, K.G.F. (2007). *Research methods in Psychology I* (PowerPoint presentation). Cape Town: University of Cape Town, Department of Psychology. Available online at:
www.ipmaac.org/mapac/meetings/1999/fall99/willmartin2.ppt
(Accessed: 2007/07/05)

[7]   Wright, B.D. & Bell, S.R. (1984). *Item banks: What, why, how?* MESA Research Memorandum Number 43, Mesa Psychometric Laboratory. Journal of Educational Measurement, 21(4): 331-345. Available online at: http://www.rasch.org/memo43.htm (Accessed: 2007/02/14)

[8]   Hay, H.R. (Ed.); Mahlomaholo, M.G.; Van der Merwe, B.C.; Brussouw, S.M.; Lampbrecht, G. & Badenhorst, J.J.C. (2004). *Manual for Teaching and Learning*. 1st Edition. Bloemfontein: Central University of Technology, Free State, Unit for Academic Development.

[9]   Merriam-Webster Incorporated. (2007). *Merriam-Webster's Online Dictionary*. Available online at: http://www.m-w.com (Accessed: 2007/06/02)

[10]   Thelwall, M. (1999). *Computer-based assessment: A versatile educational tool*. Wolverhampton: Wolverhampton University, School of Computing.

[11]   Central Michigan University. (2007). *Formative and Summative Assessment.* Available online at: http://www.provost.cmich.edu/assessment/toolkit/formativesummative.htm (Accessed: 2007/06/18)

[12]   University of Wisconsin Oshkosh College of Business. (2005). *Item Analysis*. Available online at: http://www.uwosh.edu/testing/facultyinfo/itemanalysis.php (Accessed: 2007/02/14)

[13]   Bloom, B.S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals*. Chicago, IL: Susan Fauer Company Inc., pp. 201-207.

[14]   Potgieter, P.H. (2001). *Evaluering van rekenaarpakkette as hulpmiddel by die opstel van vraestelle by die Technikon Vrystaat.* Bloemfontein: Technikon Vrystaat, Departement Inligtingstegnologie.

[15]   Valenti, S.; Cucchiarelli, A. & Panti, M. (2002). *Computer-based assessment systems evaluation via the ISO9126 Quality Model.* Journal of Information Technology Education, 1(3): 157-175.

[16]   Fulks, J. (2004). *Choosing the right assessment tools*. Bakersfield, CA: Bakersfield College.   Available online at: http://online.bakersfieldcollege.edu/courseassessment/Default.htm (Accessed: 2007/06/19)

[17]   Davis, L.   (2007). *Articulate Quizmaker: Instructional Designer ADP*. Available online at: http://www.articulate.com/products/quizmaker.php (Accessed: 2007/06/29)

[18]   Leamy, B. (2007). *ExamWeb*. Vice President: Enterprise Services, United Way of America. Available online at: http://www.examweb.com (Accessed: 2007/06/29)

[19]   Via, B. (2007). *WebQuiz*. Available online at: http://eng.smartlite.it/en2/products/webquiz/index.asp (Accessed: 2007/07/02)

[20]   Case Consulting, LLC. (2001). *CourseWebs: Learning Content Management System*. Available online at: http://www.coursewebs.com (Accessed: 2007/07/03)

[21]   Respondus. (2007). *Respondus and StudyMate*. Available online at: http://www.respondus.com (Accessed: 2007/07/24)

[22]   Wise, J.C.   (2004). *Outcomes-based assessment*. Uniontown, PA: Penn State University, Engineering Instructional Services.

[23]   European Computer Driving Licence Foundation. (2008). *Products*. Available online at: http://www.ecdl.com/products/index.jsp (Accessed 2008/07/10)

CHAPTER 2 : TEST MODELS AND SOFTWARE MEASURING INSTRUMENTS

2.1   INTRODUCTION

Chapter 1 gave a preview of computer-based assessment (CBA) and software that allows the assessor to create test items, along with a discussion of the importance of drawing up test items that successfully test whether a student is ready to be promoted to the next level. It was suggested that a test item distribution model, along with test-generating software, would assist the assessor in drawing up single- and multi-level test items. These items must be based on the curriculum outcomes, subject area, difficulty level, and Bloom's taxonomy levels.

In Chapter 2 different test models and measuring instruments are discussed. The main categories of each measuring instrument relevant to this research are also identified.

2.2   EXISTING TEST MODELS

There are different types of CBA tests, roughly grouped into five test-delivery models [1]. A test-delivery model represents the way in which test items are presented to students. On the one side of the pendulum is a test model that is not adaptive towards the performance of the student, while on the other side is a test model that is totally adaptive towards the performance of a student. This model selects a new test item from the test bank, based on whether the student answered the previous test item correctly or incorrectly. If the previous test item was answered incorrectly, the next test item that the student will receive will test the student again on the same content as the previous test item. This will continue until the student can prove that he/she has mastered that particular skill.

The five main models [1] are the Linear, Linear-on-the-Fly (LOFT), Testlet, Mastery and Adaptive models, as shown in Figure 2.1.

2.3   DISCUSSION OF MODELS

2.3.1 Linear model

In a Linear model the test items are non-adaptive, meaning that all students will receive the same test items in the same order. An advantage is that it is similar to a P&P-based questionnaire, which is familiar to most lecturers.   At the end of the test, the lecturer is able to generate a report based on the results of the group.   Another advantage is the students' ability to review, revise and omit items. One of the disadvantages, however, is the lack of security. If all students receive the same set of test items, it is possible for the paper to be revealed to those who must still write the test. If a paper is answered online, and the test items are in the same order, students are able to cheat and copy from one another.
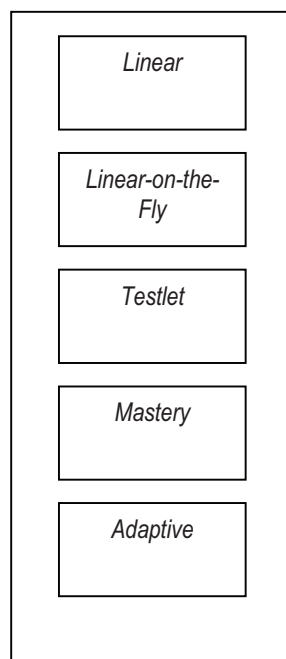
Figure 2.1: Test-delivery models

2.3.2 Linear-on-the-Fly model

With Linear-on-the-Fly tests, each student can receive different test items, but the test is of a fixed length. Randomisation can also be applied. The advantage is better test-item security, because it will be difficult for a student to determine his/her intended test items from other students who have finished writing the test, and it will also be more difficult for a student to copy from another student seated nearby. Students will still be able to review, revise and omit items. Since each student has a different test, there must be a large pool of items from which test items can be selected. Test items are drawn from the same databank.

2.3.3 Testlet model

The Testlet model involves a group of test items related to a specific content area. Items are grouped together based on (a) difficulty level, (b) subject matter, and (c) single- and multi-tasking. Multi-tasking refers to test items where the student must complete more than one action when answering a question. A student will receive different test items in a test, but of equal difficulty and based on the same content. It will be easier to ensure security, because students will receive different tests.

2.3.4 Mastery model

With the Mastery model it is possible to determine whether a student has mastered certain skills, based on decision rules in the database. If a student experiences difficulty in answering test items from a specific subject area, the student will keep on receiving test items from that area until he/she has mastered the content. An advantage of this model is its efficient nature, because a student is "classified based on simple decision rules" [1]. This model is not used, because it is a computer-based training model rather than a computer-based assessment model.

2.3.5 Adaptive model

The Adaptive test is a well-known model in use today. The student receives test items based on his/her performance. Each student receives different test items based on his/her responses while writing the test. The test can be of a fixed length or ongoing until the student reaches a proficiency scoring level. The Item Response (IR) theory uses a method where items have a specific difficulty level and are presented to the student according to his/her abilities at that stage. Security is one advantage of this model, another being the limitation in test length.   If it is clear that a student has mastered certain skills, the test is terminated.

2.4   MODEL USED IN THIS STUDY

For purposes of this study, a combination of the Linear-on-the-Fly (LOFT) model and the Testlet model is used. The importance of the LOFT model lies in its ability to:

- o Deliver different tests of the same length to students;
- o Randomise test items for security purposes; and
- o Allow the students to omit, review and revise test items.

The Testlet model ensures that a question paper consists of:

- o Test items that are related to a specific content area;
- o Items that are equal in difficulty, where most of the test items are multi-task items; and
- o Test items based on the same difficulty level.

The standard method used at the CUT to assess students in Microsoft Access is the P&P-based test. Students are divided into groups, and the students in each group receive the same paper. They then complete the test on computer. A

single computer lab can accommodate about 50 students at a time. An attempt is made to give the different groups different papers, but of the same length and with test items equal in difficulty level. In order to minimise student cheating within a single group, it is recommended that test items be randomised. This simulates a typical LOFT model.

Test items in Microsoft Access are related to different content areas, which are again related to the curriculum outcomes of the subject. These test items are mostly multi-task test items. An example of a multi-task test item is where a student is asked to create a table, which will require the student to perform several actions. The student will have to create fields, add attributes to the fields, and save the table. This is typical of a Testlet model.

2.5   EVALUATING EXISTING MEASURING INSTRUMENTS

There is a variety of assessment tools that lecturers can use to draw up tests. Some packages, such as Respondus [2], allow the lecturer to create test items from the start and store the test items in an empty databank. Other packages, such as compAssess [3], allow the lecturer to use test items from an existing test bank or to create test items and store the items in the same databank. There are measuring instruments in place to validate these assessment tools.

2.5.1 Measuring Instrument 1

The first measuring instrument under discussion was developed by Potgieter [4] to evaluate computer packages that allow the user to draw up papers and memoranda. Potgieter's research revealed an overall need amongst lecturers to use CBA software that would enable them to draw up papers and memoranda where numbering is automatically applied. They also expressed a need to be able to search for certain words and replace them with others, to perform spellchecking, and to have access to WYSIWIG (What You See Is What You

Get) – a feature where the newly created paper on the computer compares exactly with the final printed version.

Other needs expressed by assessors were the ability to apply randomisation, and also to export the paper and memorandum to a well-known program like Word in Microsoft Office, where advanced editing can take place, which would allow the assessor to make changes to the content or structure of the paper after it has been constructed. Assessors also identified a need to print papers from the program where the test items were generated or to export the papers to another program from where they can be printed.

The Help function is also a useful guide for assessors in drawing up a paper. The assessors interviewed for Potgieter's research requested a variety of question types, like True/False, Short-Answer and Multiple-Choice. Moreover, being able to classify a test item according to its difficulty level would help the assessor to draw up different papers with the same difficulty level. The study revealed that assessors tended to avoid classifying items in terms of difficulty, which is sometimes subject to a student first writing a test in order to determine the difficulty level of such an item. When using test items from a databank, the assessor should be able to Search (Find) for test items of a certain difficulty level, as well as test items linked to a specific subject matter.

Some of the main categories in Potgieter's instrument can be viewed in Table 2.1 below.

Table 2.1: Evaluation categories for computer packages

Category	Description
Formatting of text items	Formatting/editing of text in program; Numbering; Finding and replacing words; Running Spellchecker; WYSIWYG
Paper and memorandum construction	Registering each test item and its answer; Adding/deleting any extra information on test structure; Randomly selecting a fixed number of questions from the databank
Presentation method: Printing of paper and memorandum	Applying numbering with the software (no need to export a document to another software program to apply numbering); Printing the paper directly from the program or presenting the questions online
Exporting of paper/ memorandum	Exporting the test to a program like Word in order to make changes
Help	Basic guidance where needed
Find	Finding test items with a certain difficulty level or of a specific subject matter
Item classification	Availability of different question types like True/False and Multiple-Choice; Sufficient space to answer paragraph-type questions; Place to show percentage rating of knowledge/insight/application/ingenuity of paper; Difficulty level of each question

2.5.2 Measuring Instrument 2

Fulks [5] developed a second assessment-grading tool in the form of a checklist divided into six categories (Table 2.2).   For purposes of this study, only the question types applicable to Microsoft Access have been included in the table. Fulks identified certain advantages and disadvantages of the different assessment types.

In Fulks' checklist, multiple-choice test items are items where a range of possible answers is given to a student. The student must then select one or more correct answers. These test items are easy to grade, but need careful construction.

Licensing examinations are drawn up from test items previously evaluated for validity. Fulks [5] mentioned that licensing examinations have the advantage of analysing the validity and reliability of the test items beforehand. The difficulty level of test items can be determined in advance, which makes it easy for the lecturer to decide which test items to use in a test.

It is possible that the quality of a test item is below standard, improperly evaluated, or even outdated. The student's test-writing skills can also have an influence on his/her performance.   Ultimately, there are numerous factors that make it difficult to test a student's true ability.

Case studies, as described by Fulks, allow a student to think analytically and display synthetic thinking skills. Synthetic thinking skills involve planning, building and developing, which form the essence of a design process.   This allows students to incorporate knowledge from other areas (subjects) or disciplines, especially where group work can be done.   It takes a lot of time to develop case studies. In order to prevent project creep, students must have well-defined criteria.

Test items that involve problem-solving (Table 2.2) cater for analytical and synthetic skills to be evaluated in real-world situations. This becomes difficult to assess, because there are different methods that can be used to solve problems.

Table 2.2: Assessment grading tool

Type of question	Domain: C= Cognitive A= Affective C/A= Cognitive or Affective	Usage Type: F= Formative S= Summative F/S= Formative or Summative	Advantages	Disadvantages
Multiple-choice	C	F/S	Easy to grade, objective	Reduces assessment to multiple-choice answers
Licensing exams	C	S	Easy to score and compare	No authentic testing, may be outdated
Case study	C/A	F/S	Efficiently displays analytical and synthetic thinking; connects other knowledge to topic	Creating the case is time-consuming; dependent on student knowledge from multiple areas
Problem-solving	C	F or S	Efficiently displays analytical and synthetic thinking; authentic if real-world situations are used	Difficult to grade due to multiple methods and potential multiple solutions

The student may not be led by the options from which he/she may select an answer. These types of test items do not adequately test higher level knowledge, but instead lie on the cognitive level. This is also referred to as Bloom's taxonomy (discussed in Chapter 3).

Fulks' [5] assessment and grading tool contains a field known as the Cognitive and Affective Domain. Cognitive skills help a person to apply knowledge, which is then used to solve problems. Affective skills address a person's emotions in a specific learning experience.

The usage type in Fulks' table is defined as Formative or Summative. Summative assessment is carried out at the end of a course/project, while formative assessment is ongoing throughout the course.

2.5.3 Measuring Instrument 3

2.5.3.1 Introduction

In 1991 the International Organization for Standardization [6] defined the ISO9126 standard for Information Technology (Software Quality Characteristics and Sub-characteristics), which identifies six quality characteristics according to which a CBA system must be evaluated, namely functionality, usability, reliability, efficiency, portability, and maintainability. For each characteristic there are also certain sub-characteristics (Table 2.3). Each sub-characteristic is further divided into attributes, which are entities that can be measured in a software product [7].

To test the reliability of a system, a system has to be evaluated under different conditions to try and determine the frequency of certain failures. Usability is difficult to measure, because objectiveness is influenced by personal perceptions. Efficiency is determined by the availability of resources within a company.

Table 2.3: Characteristics according to which a CBA system can be evaluated

Characteristic	Sub-characteristics	Explanation
Functionality	Suitability	Does the TMS/TDS fit its purpose?
	Accuracy	Accuracy of results.
	Interoperability	Working together with other programs.
	Compliance	According to specific standards/laws?
	Security	Is security trustworthy?
Usability	Learnability	Is the system easy to learn?
	Operability	Is the system easy to operate & control?
	Understandability	Is the system easy to understand?
Reliability	Maturity	How often does the system fail?
	Recoverability	How easily can the system recover from failure?
	Fault tolerance	Can one still perform basics with the system after failure?
Efficiency	Time behaviour	Response time, processing time & throughput rates.
	Resource behaviour	Quantity of resources used.
Portability	Installability	Easy to install?
	Replaceability	Can it replace other software in that environment?
	Adaptability	Adaptability of software to function in other environments and with other software.
Maintainability	Stability	How stable is the software in the event of modification?
	Analysability	How easy is it to diagnose deficiencies?
	Changeability	How easily can software be modified?
	Testability	How easy is it to retest software after modification?

Portability is partially dependent on the domain where it is installed.   To maintain a system, insight is needed into the coding, which makes it very difficult for the user to evaluate.   Because the focus is mainly on the educational task, this dissertation pays attention to the functionality of a system.

2.5.3.1.1 Functionality

Functionality is divided into certain sub-characteristics, namely Suitability, Accuracy, Interoperability, Compliance and Security.

Functionality tests whether a system is suitable for its purpose and whether the software is in harmony with the system, as well as the trustworthiness of the security and whether the data, people and systems are connecting.

2.5.3.1.2 CBA Components

A typical CBA system (Figure 2.2) consists of the following:

- o A Test-Management System (TMS), consisting of a user interface, which allows the assessor to create test items and evaluate students' answers;
- o A Test-Delivery System (TDS), which delivers tests to students over the Web, on LAN or on individual computers (Table 2.4).

Valenti, Cucchiarelli and Panti [8] took computer-based assessment systems and divided them into a TMS and a TDS.   The TMS was divided into two areas, namely the Question-Management Area (QMA) and the Test-Management Area (TMA).   The TDS was also divided into a QMA and a TMA (Table 2.4).

2.5.3.1.2.1 Suitability of the Test-Management System (TMS)

2.5.3.1.2.1.1 Question-Management Area (QMA)

The QMA focuses on different question types and their structures. There are many question types in CBA tests, but the basic types are True/False, Multiple-Choice, Short-Answer, and Fill-in-the-Blank questions.

2.5.3.1.2.1.2 Test-Management Area (TMA)

In the TMA, test items are selected and organised into tests. A lecturer is able to select test items for a test, based on topic, difficulty level, curriculum outcome, cognitive level and many other attributes. If randomisation is available it can be applied. An attribute is an entity that can be measured in a software product [7].

To test the reliability of a system, it must be evaluated under various conditions to try and determine the frequency of certain failures. Usability is difficult to measure, because objectiveness is influenced by personal perceptions. Efficiency is determined by the amount of resources available in a company. Portability is partially dependent on the domain where it is installed. To maintain a system, insight is needed into the coding, which makes it very difficult for the user to evaluate. Because the focus is mainly on the educational task, this dissertation pays attention to the functionality of a system.

When a test is assessed, the instructor can expect to get certain feedback:
- o The percentage rating of the individual's performance;
- o A summary of the individual's item responses;
- o A summary of the performance of the group, showing the distribution, means and deviations; and

- An analysis of the statistics of each item after a test has been written and from which information on the reliability, discrimination and difficulty of a test can be drawn.

2.5.3.1.2.2 Suitability of the Test-Delivery System (TDS)

2.5.3.1.2.2.1 Question-Management Area (QMA)

The QMA focuses on:
- The number of attempts a student is allowed to answer a test item;
- The time duration scheduled for the test; and
- Feedback in the form of sketches and other available multimedia to guide a student.

Test-Management System (TMS)

Test-Delivery System (TDS)

Format

Apply features like:
Curriculum Outcome
Question Type
Question Level
Randomisation
Subject Area
Thinking Skills

Apply formatting like:
Fonts
Numbering
Spellchecking
Find/Replace
Exporting for Editing

Databank

Questions

Form Designer

Publish

Online availability and student responses

Assessment

Test-Assessment Results

Reports

Student Results

Question Evaluation

Section Evaluation

Figure 2.2: Test-Management System and Test-Delivery System

Table 2.4: Question-Management Area and Test-Management Area of a Test-Delivery System

		TMS	TDS
Functionality	Suitability	QMA (question types + structure) TMA (test bank, test preparation, assessment, statistics)	QMA (retries, feedback, multimedia) TMA (databank/student responses)
	Security		

2.5.3.1.2.2.2 Test-Management Area (TMA)

A CBA test normally involves the assessment of the test directly after the student has completed it. The student's responses are saved in the TMA and he/she receives his/her marks once the Finish button has been clicked. This information must be saved on the system for the lecturer, who is normally also the system administrator.

2.5.3.1.2.3 Security

Security is of high importance. Students may not have access to a databank or tests if not made available to them, or to the personal data of any other students. Restricted availability will ensure security. Time restrictions on tests can be applied where a test is made available to a specific group during a specified time of the day. This minimises the chances of another group accessing the same test beforehand.

2.6 CHARACTERISTICS ACCORDING TO WHICH PACKAGES ARE EVALUATED

Three measuring instruments have been discussed, and items from all three are used to evaluate existing software packages. Each software package is evaluated on the basis of its ability to help the assessor draw up test items. Based on the instrument used by Potgieter, the items Formatting, Presenter, Randomisation, Help, and Question Types are used. The second instrument, as developed by Fulks, focuses on Bloom's Taxonomy levels linked to different question types. From the third instrument, as described by the ISO9126 standard, Functionality with the sub-characteristic Suitability is used.

A few extra attributes are added to the list in order to evaluate existing assessment software packages. The Testlet model focuses on test items that are related to a specific content area. The term 'Subject Area' is then added when evaluating existing software packages and replaces the term 'Content Area'. The Subject Area references the main content of each learning module (Chapter 3). Each test item is related to a subject area such as Creating Tables and Queries, Printing and others. The difficulty level of a test item is also included.

The importance of curriculum outcomes and multi-level test items was mentioned in Chapter 1. Curriculum outcomes and multi-level test items are also used in the evaluation of existing software packages. The items used from the three instruments are grouped in

Table 2.5: Characteristics according to which software packages that create test items can be evaluated

Criteria	Details
Formatting	Numbering Find/Replace Spellchecking Exporting for editing
Presentation	Presenting a test online Giving students a test as a printout
Randomisation	Arranging items
Help	Guidance received in drawing up a paper
Question Types	Multiple-Choice, True/False, Case Study, Licensing Examinations, Problem-Solving, etc.
Software Functionality	Checking how well/easily software is working and the amount of resources used
Bloom	Bloom's taxonomy levels
Subject Area	Linking items to subject area
Curriculum Outcomes	Matching test items with outcomes in a curriculum
Multi-Level Test Items	Level 1, Level 2, Level 3

2.7 CONCLUSION

Chapter 2 discussed different test-delivery models, namely the Linear, Linear-on-the-Fly (LOFT), Testlet, Mastery, and Adaptive test models. For purposes of this study, the LOFT and Testlet models are used.

There are different measuring instruments in place to validate assessment tools, which allow the lecturer to draw up test items and to assess students online.

Chapter 3 discusses the curriculum outcomes for Microsoft Access, as offered at the CUT, as well as Bloom's cognitive levels.

2.8 REFERENCES

[1]   Patelis, T. (2000).   An overview of computer-based testing. *Research Notes, RN-09*. New York: College Board.

[2]   Respondus. (2008). *Respondus 3.5*. Available online at: http://www.respondus.com/download/respondus.shtml (Accessed: 2008/01/15)

[3]   SVI Training Products, Inc. (2008). *compAsses Skills Assessment Program*. Available online at: http://www.svitrain.com (Accessed: 2008/04/08)

[4]   Potgieter, P.H. (2001). Evaluering van rekenaarpakkette as hulpmiddel by die opstel van vraestelle by die Technikon Vrystaat.   Bloemfontein: Technikon Vrystaat, Departement Inligtingstegnologie.

[5]   Fulks, J. (2004).   *Assessing student learning in community colleges*. Bakersfield, CA: Bakersfield College. Available online at: http://online.bakersfieldcollege.edu/courseassessment/Default.htm (Accessed: 2008/04/08)

[6]   International Organization for Standardization. (1991). *Information Technology – Software quality characteristics and sub-characteristics: ISO/IEC 9126-1*. Geneva: ISO.

[7]   Wikipedia. (2007). *ISO 9126*. Available online at: http://en.wikipedia.org/wiki/ISO_9126 (Accessed: 2008/04/12)

[8]   Valenti, S.; Cucchiarelli, A. & Panti, M. (2002).   Computer-based assessment systems evaluation via the ISO9126 Quality Model. *Journal of Information Technology Education*, 1(3): 157-175.

CHAPTER 3 : BLOOM'S TAXONOMY LEVELS AND CURRICULUM OUTCOMES

3.1 INTRODUCTION

Test items must successfully test whether a student is ready to be promoted to the next level of study. It has been suggested that a test-item distribution model, along with test-generating software, would help the assessor in drawing up single- and multi-level test items.

Chapter 2 took a closer look at available models that can be used to evaluate CBA packages. Three measuring instruments were discussed and a table was created with criteria to be used in Chapter 4 to evaluate test-generating software.

This chapter, Chapter 3, discusses the following:

- Bloom's taxonomy levels;
- Relevant verbs that can be linked to Bloom's relevant cognitive levels; and
- The current distribution of test items on Bloom's taxonomy levels for first-, second- and third-year students taking Microsoft Access at the CUT;
- Modules, curriculum outcomes and the subject areas/object types for Microsoft Access presented in the subject INL20DB for IT students at the CUT.

3.2 BLOOM'S TAXONOMY LEVELS AND RELEVANT VERBS

In the 1950s, Benjamin Bloom and a team of educational psychologists analysed different academic learning behaviours [1]. The model that was subsequently developed is known as Bloom's taxonomy, and is still in use in classrooms today. This model categorises learning behaviours into three main domains, namely the cognitive (knowledge), affective (attitude) and psychomotor (skills) domains. This study focuses on the cognitive domain, which represents a person's intellectual abilities. During the 1990s, improvements were made to Bloom's taxonomy levels to make them more relevant in the twenty-first century [2].

In Table 3.1 the cognitive domain is categorised into six different skills, arranged from the simplest to the most difficult skill. The first level is the knowledge level, where students must recognise information and be able to recall data. An example of a test item on the knowledge level would be to ask a student to name the different steps in normalisation.

The second level, which evaluates the comprehension skills of the student, builds on the knowledge level, meaning that a student must possess certain knowledge in order to answer related test items on the comprehension level. Test items on this level test whether a student has the ability to understand and explain work content. The CUT's Manual for Teaching and Learning, edited by Hay [3], mentions that this includes the translation of data that is in a specific form, like a graph, to another form such as writing. An example of a test item at this level would be to ask a student to communicate the results of a query.

The third level is the application level, and the student requires knowledge and "a degree of comprehension" [3] to answer test items on this level. Knowledge must be applied in a familiar situation in a manner that will result in change. An

example of a test item on the application level would be to ask a student to modify an object in a database.

On level four the student must apply certain analysis skills and break down current data into more understandable units. All three foregoing levels, namely the knowledge, comprehension and application levels, will be present, and the student may be asked to analyse a fact or a relationship. In a database it is important to analyse data before tables can be created. A key part of database structure is the normalisation [4] of data – a technique used to minimise the duplication of information in order to safeguard a database against logical and structural problems. An example of such a test item would be to ask a student to normalise tables.

At the top of this domain, in categories five and six, the evaluation and synthesis levels are found. Which category should occupy the top position is debatable [3]. The evaluation and synthesis levels involve all of the aforementioned skills levels. Skills on the evaluation level are applied by learners when they can form their own opinions. The student will have the ability to judge information and make recommendations. An example of a test item on the evaluation level would be to ask the student to validate an existing Microsoft Access database and to make recommendations on ways to improve the database.

On the synthesis level, students can creatively combine existing information into something new. This skill will allow a student to create new tables, queries, forms, reports, pages, macros and modules from existing information. A typical test item representing this skill would be to ask a student to create a new object in a database.

Table 3.1: Cognitive domain with skills levels

Level	Definition	Verbs	Questions
Knowledge (The student must <u>remember</u> or recall/retell data)	The student memorises data and is then able to recall or retell the data.	define, draw, describe, identify, label, list, match, name, recall, recognise, repeat, retrieve, select, state, underline/circle, write	1. Name any four objects in a database.
Comprehension (The student must <u>understand the data he/she is working with</u>)	The student understands and interprets data based on prior learning.	associate, defend, discriminate, discuss, explain, express, give in own words, group, interpret, paraphrase, report, restate, review, summarise, tell, translate	1. Explain why you have normalised the tables as you did. 2. Explain the results of your query.
Application (The student must be able to <u>apply</u> his/her new knowledge)	The student applies his/her knowledge in a familiar situation.	apply, calculate, carry out, compute, construct, demonstrate, implement, modify, prepare, solve, use, utilise	1. Use an existing table and add/ modify a field. 2. Open an existing query and calculate hours worked per week of each worker.

Analysis (The student must be able to <u>analyse</u> data)	The student interprets knowledge in order to apply principles and relationships in a situation.	analyse, compare, contrast, detect, determine, develop, diagnose, draw, estimate, examine, identify, inventory, organise, predict, separate, solve, test	1. Create tables and build relationships between the tables. 2. Apply normalisation. 3. Compare the results of the two queries with each other.
Evaluation (The student must <u>evaluate</u> data)	The student develops his/her own opinions and is able to judge others' opinions and decisions on the basis of certain standards.	appraise, assess, check, choose, critique, compare and evaluate in order to make judgments**,** estimate, how would you, judge and recommend, justify, measure, rate, revise, score, select, test, validate, value, what do you think	1. Validate an existing database and make recommendations to improve the database. 2. Test a query. 3. Make recommendations on table structures.

Synthesis (The student must <u>create</u> something new)	The student is able to use old ideas to create something new.	arrange, create, collect, combine, compose, construct, design, develop, manage, organise, plan and prepare, produce, set up	1. Create a new query using fields from an existing table. 2. Design a switchboard that will allow one to move to queries and forms. 3. Combine the information in table 1 and table 2 and create a report.

Microsoft Access is the practical component of the subject INL20DB.   In order to answer test items that lie on the application, analysis or synthesis level, students need knowledge and comprehension. Test items on the evaluation level will test the student's skill in explaining or describing something. Due to the practical nature of Microsoft Access, the focus of this study is on the application, analysis and synthesis levels.

3.3   DISTRIBUTION OF TEST ITEMS AND BLOOM'S TAXONOMY LEVELS

INL20DB students doing Microsoft Access must be able to develop databases. Although it is important for the students to remember, understand and evaluate data, they will not be evaluated on test items directly linked to the knowledge, comprehension and evaluation levels. The focus will be on test items that the assessor can assess in a database without additional explanation from the student's side. Typical verbs used in test items on the knowledge,

comprehension and evaluation levels, in terms of information, are "list", "describe", "interpret", "summarise", "judge" and "critique".

According to Bocij and Greasley [5] the majority of assessment packages available seem incapable of testing or developing cognitive skills at the higher levels of analysis, synthesis and evaluation.   There is no set rule that determines the percentage of test items on each of Bloom's taxonomy levels [3]. The CUT's Manual for Teaching and Learning, edited by Hay [3], suggests activity distribution (Table 3.2), where the higher level activities are found more often in senior course years.

Table 3.2: Test-item distribution

	Year 1	Year 2	Year 3
Knowledge	40	30	10
Comprehension	20	20	10
Application	20	20	10
Analysis	10	10	20
Synthesis	10	10	20
Evaluation	00	10	30
Total	100	100	100

Table 3.3 reflects a rough estimation of the distribution of test items on the synthesis, application and analysis levels for first-, second- and third-year students in the subjects INL10DB, INL20DB and INL30EB. These students created databases in Microsoft Access and the lecturers teaching these students were asked to roughly indicate the distribution of Microsoft Access test items on the application, synthesis and analysis levels when drawing up tests.

Table 3.3: Test-item distribution in Microsoft Access

	Year 1	Year 2	Year 3
Application	60	30	10
Synthesis	30	50	40
Analysis	10	20	50
Total	100	100	100

First-year INL10DB students at the CUT are assessed by writing online ECDL (European Computer Driving Licence) [6] Microsoft Access tests. Although the students learn to create objects, the online tests mainly test students on the application level. According to the assessor interviewed for purposes of this study, roughly sixty percent of the test items lie on the application level, thirty percent on the synthesis level, and only ten percent on the analysis level. The test items are drawn from an online test bank. The assessor does not have access to these test banks and can therefore not determine the distribution of test items on the application, synthesis and analysis levels. If an assessor uses ECDL, then he/she must also follow the curriculum of ECDL.

Second-year students receive paper-based tests, and they are expected to create new objects or change existing objects on the computer. The assessor interviewed roughly estimated that thirty percent of the test items lie on the application level, fifty percent on the synthesis level, and twenty percent on the analysis level. In the second year the focus shifts from application-level test items to test items on the synthesis level.

Third-year students receive paper-based assignments that they complete on computer. These students are required to analyse data and normalise tables in order to create databases. A rough estimation, according to the assessor interviewed, is that ten percent of the test items lie on the application level, forty percent on the synthesis level, and fifty percent on the analysis level.

Chapter 6 takes a closer look at IT students' performance in Microsoft Access as part of the subject INL20DB, specifically on the application and synthesis levels. An attempt is made to recreate a test-item distribution model, which will suggest how test items in Microsoft Access should be distributed on the application, analysis and synthesis levels for second-year students.

3.4 MODULES, CURRICULUM OUTCOMES AND SUBJECT AREAS/ OBJECT TYPES

The term *curriculum* refers to the content of a particular course as determined by an authoritative body. Students choose a range of subjects from a particular course, and the curriculum of each subject includes the "teaching, learning and assessment materials" applicable to that subject in the course [7].

The objectives of a subject can be expressed as learning outcomes and assessment strategies. These outcomes and assessments are grouped into modules, which in turn are the building-blocks of a qualification [3]. A module comprises a specific part of the curriculum. The curriculum for second-year Information Technology students at the CUT stipulates a thorough knowledge of Microsoft Access 2002.

Each module is divided into assessment criteria, which are used in defining the curriculum outcomes and subject areas/object types of each module. The outcomes can be referred to as exit-level outcomes [3]. Through assessment it is possible to determine a student's level of knowledge and skill in a specific subject [8]. The subject area/object type covered in a specific module contains objects (tables, queries, forms, etc.) and main tasks covered in the module. In Table 3.4 the curriculum content of Microsoft Access for INL20DB students is divided into ten tutorials, with a breakdown of the curriculum outcomes in the second column. In the third column, each tutorial is assigned keywords

identifying possible subject areas/object types in that specific tutorial. For instance, the curriculum outcomes of Tutorial 1 will fall in the subject area/object type "TABLE" or "DATABASE". If any changes occur within a tutorial, changes can be made to the curriculum outcomes and the subject area/object type.

Table 3.4: Assessment criteria in Microsoft Access

	ASSESSMENT CRITERIA	
Modules	Curriculum Outcomes	Subject Area/ Object Type
Module 1 Introduction to Microsoft Access 2002	After this tutorial, learners will be able to: Define database concepts; Start and exit Access ; Open an existing database; Learn how Access saves a database; Identify the components of the Access and Database windows; Open, navigate, and print a table; Open, navigate, run and print a query; Create a query; Create a form; Navigate, preview and print a form; Use the Access Help system; Create a report; Preview and print a report; Manage databases by backing up, restoring, compacting, repairing and converting databases.	DATABASE TABLE
Module 2 Maintaining a Database	After this tutorial, learners will be able to: Understand the guidelines for designing databases; Create a new database; Create and save a table; Define fields for a table and specify the primary key; Set field properties; Add records to a table; Delete, move and add fields; Change field properties;	DATABASE TABLE

	Copy records from another Access database; Delete and change records; Import tables from another Access database; Define a one-to-many relationship between two tables.	
Module 3 Querying a Database	After this tutorial, learners will be able to: Create, run, and save queries; Update data using a query; Define a relationship between two tables; Sort data in a query; Filter data in a query; Specify an exact match condition in a query; Change a datasheet's appearance; Use a comparison operator to match a range of values; Use the And and Or logical operators; Perform calculations in a query using calculated fields, aggregate functions, and record group calculations.	QUERY
Module 4 Creating Forms and Reports	After this tutorial, learners will be able to: Create a form using the Form Wizard; Change a form's AutoFormat; Navigate a form and find data using a form; Preview and print selected form records; Maintain table data using a form; Check the spelling of table data using a form; Create a form with a main form and a sub-form; Create a report using the Report Wizard; Insert a picture in a report; Add, move, resize and align controls in a report; Modify control properties; Add a sub-report to a main report;	FORM REPORT

	Add lines to a report; Add calculated controls to a report; Calculate group and overall totals in a report; Define conditional formatting rules; Use domain aggregate functions; Add the date, page, numbers and title to a report; Create and modify mailing labels; Preview and print a report.	
Module 5 Creating More Advanced Queries and Custom Forms	After this tutorial, learners will be able to: Create a Lookup Wizard field in a table; Display related table records in a sub-datasheet; Create an input mask for a table field; Define multiple selection criteria in a query; Specify data validation values; Use the Like and Not operators in a query (pattern match); Use the In operator in a query (list-of-values); Use both the And and Or logical operators in the same query; Create a parameter query; Design and create a custom form; Select, move, and delete controls; Add form headers and footers; Add a picture to a form; Use Control Wizard to create a multi-page form; Use a filter to select and sort records in a form.	FORM TABLE QUERY FORM
Module 6 Customising Reports and Integrating Access with other Programs	After this tutorial, learners will be able to: Design and create a custom report; Assign a conditional value to a calculated field; Modify report controls and properties; Sort and group data; Calculate group and overall totals; Hide duplicate values;	REPORT

	Embed and link objects in a report; Export Access data to other programs.	
Module 7 Working with HTML Documents, Data Access Pages, and Hyperlink Fields	After this tutorial, learners will be able to: Export an Access table to an HTML document; View an HTML document using a browser; Use a Wizard to create a Data Access page for an Access table; Update a Data Access page using a Web browser; Sort and filter Data Access page records; Create a custom Data Access page; Import an HTML document as an Access table; Add hyperlink fields to an Access table; Create hyperlinks to Office documents and Web pages.	PAGES
Module 8 Using Query Wizards, Action Queries, and Replication	After this tutorial, learners will be able to: Use Query Wizards to create a Cross-tab query, a Find Duplicates query, and a Find Unmatched query; Create a Top Values query; Create an Action query; Define many-to-many and one-to-one relationships between tables; View and create indexes for tables; Join a table using a self-join; View SQL query statements; Use replication to create a Design Master and Replica of a database; Synchronise the Design Master and Replica databases.	QUERY REPLICATION

Module 9	After this tutorial, learners will be able to:	
Automating Tasks with Macros	Design a switchboard and dialog box for a graphical user interface; Run and add actions to macros; Single-step a macro; Create a macro; Add a macro to a macro group; Add a command button to a form; Attach a macro to a command button; Define data validation criteria; Create a dialog box; Add a list box to a form; Use an SQL statement to fill a list box; Create a macro group; Use the Switchboard Manager to create a switchboard.	MACRO SWITCHBOARD
Module 10	After this tutorial, learners will be able to:	
Using and Writing Visual Basic for Applications Code	Review and modify an existing sub-procedure in an event procedure; Create function procedures in a standard module; Create event procedures; Compile and test function procedures, sub-procedures and event procedures; Hide text and change display colours; Encrypt and decrypt a database; Set and unset a database password; Analyse a database; Split a database; Set database start-up options.	DATABASE MODULES

The tutorials build on one another from Tutorial 1 (introduction level) to Tutorial 10 (advanced level).

3.5 LINK BETWEEN BLOOM'S TAXONOMY LEVELS AND THE CURRICULUM

On the application level the student must apply his/her knowledge to existing objects like tables and queries. The student can, for example, be asked to modify a table by adding a field to the table or by doing a calculation in a query.

Currently it seems like the majority of test items for second-year IT students in the subject INL20DB are on the synthesis level. On this level students must design new database objects like tables and queries.

On the analysis level the student must interpret knowledge in order to apply principles and relationships. Third-year students currently focus on test items on the analysis level. The student must understand the data in order to build meaningful relationships between table objects. In this category, data outcomes are analysed and compared. A student may be asked to develop a new database and to normalise the tables in the database.

Table 3.5 gives a list of verbs that can be used to draw up Microsoft Access test items for INL20DB students on Bloom's application, analysis and synthesis levels.

It is possible that the same verb can be used for tasks lying on different levels. A student can be asked to *create* a relationship (analysis level) or to *create* a new table (synthesis level).

Table 3.5: Verbs used in drawing up test items in Microsoft Access

Bloom's Level	Verbs
Application	Add, align, apply, assign, attach, back up, calculate, change, compact, compile, convert, copy, create, decrypt, delete, embed, encrypt, export, filter, hide, import, insert, join, link, modify, move, navigate, open
Synthesis	Calculate, create, design, display, embed, perform, preview, print, repair, resize, review, run, save, select, set, sort, split, synchronise, unset, update, use, view
Analysis	Analyse, define

3.6   MULTI-LEVEL TEST ITEMS

Like modules, test items can also build on one another. For purposes of this study, the term "multi-level test items" is used to refer to test items that build on one another (Chapter 2, paragraph 2.4).

3.6.1 Determining the level of a test item

A database can consist of various objects, namely tables, queries, forms, reports, pages, macros and modules (Figure 3.1). The main object in a database is a table. Once a table has been created, the database developer can create other objects like queries, forms, reports, pages, macros and/or modules.

In this discussion, the level of a test item is based on the scenario that the student had to create a database from scratch. The level of a test item is determined on the basis of Table 3.6, which consists of certain fields, namely S1 (single-level), M2 (multi-level 2), M3 (multi-level 3) and M4 (multi-level 4). There can be any number of fields. The test that was given to the students consisted of twelve main test items, some of which were subdivided into smaller sets of test

items.   In total the students were evaluated on fifty-eight actions they were required to complete.   The different test items given to the INL20DB students (Appendix A:A1) were arranged in rows, as shown in Table 3.6.

Newly created tables will normally lie on S1, because they do not depend on any previously created objects.   There are situations where tables can be created through queries, which will place such an item on another level. The students received a test requiring them to create the 'AthleteEvent' table (Appendix A:A1). This table could be created without the need to create any prior object and is therefore an S1 test item (see Test Item 1 in Table 3.6).   When the level is determined, the crossing point between the level and the test item is marked with the letter "X".
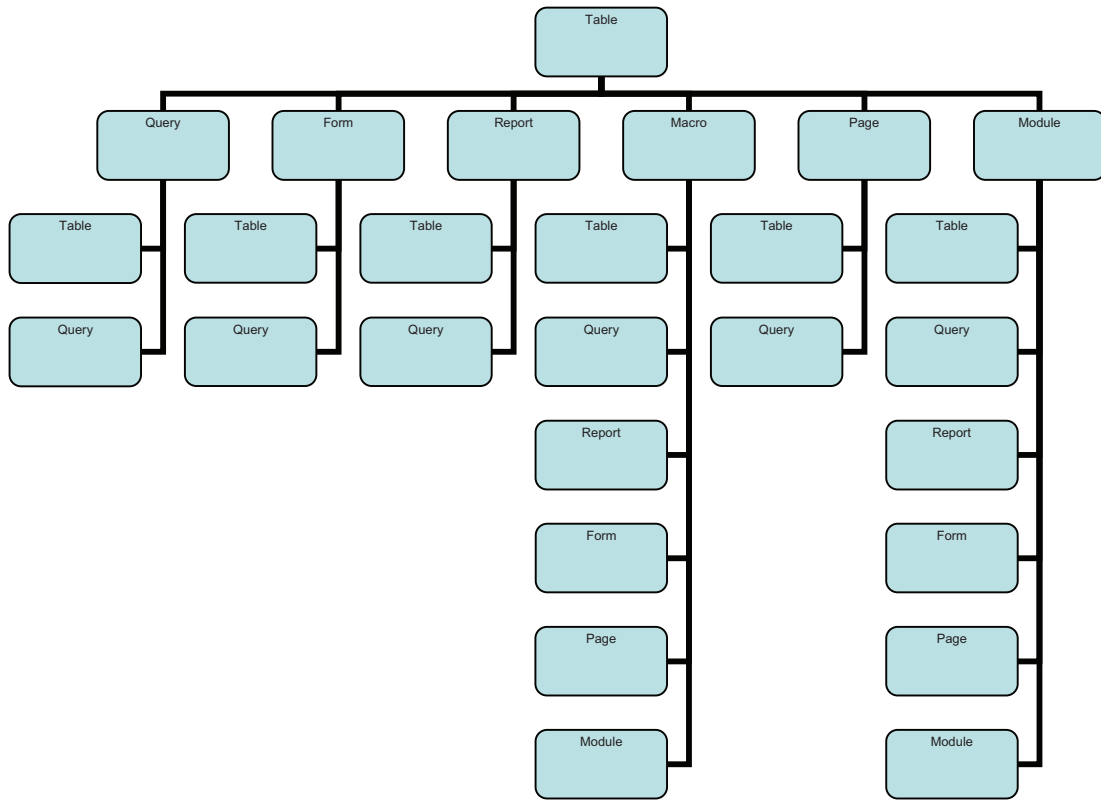
Figure 3.1: Microsoft Access objects

Queries can be categorised as M2 test items. There are situations where queries can be S1 test items, especially where tables are provided. In order to create a query, information from another query or table must be used (Figure 3.1). Students were asked to create the query "Age25+" (see Test Item 10.1, Table 3.6). Since the students first had to create a table (S1), the query was now dependent on the table created previously, which makes it an M2 test item.

Forms will normally be M2 or M3 test items. Forms are created from tables (S1) and/or queries (M2). The students were asked to create a switchboard (also called a form) that opened the query "Age25" (Test Item 11.2). Before they could create the switchboard, they had to create a table (S1), a query (M2) and a macro (M3). This makes the switchboard item an M4 test item.

Macros are normally M2, M3 or M4 test items. Macros can perform actions on any object in the database. Students were asked to create a macro that would open the query "Age25+" (Test item 11.2). Before the students could create the macro, they had to create a table (S1) and a query (M2). The macro is therefore an M3 test item.

There were no test items on reports. Reports can be assigned to the same category as forms. A report is normally created after tables and queries have been created.

It is possible that one assessor may differ from another in determining the level of a test item. The assessor must feel free to use his/her discretion and interpretation skills in determining the level of a test item.

Table 3.6 reflects seventeen S1 test items, six M2, four M3 and three M4 test items. The development of the software product as reported on in Chapter five will allow any number of levels and allow the assessor to indicate the level on which a test item lies.

Table 3.6: Determining the level of a test item

Test Item	Test Item Content	Level 1	Level 2	Level 3	Level 4
1	Open the database SportEvent1				
2	Create AthleteEvent1	X			
3	Add data	X			
4	Compact on Close	X			
5	Relationships		X		
6	Query 1: Age25+		X		
7	Query 2: Bfn25Sprints		X		
8	Query 3: DHSItems		X		
9	Query 4: DHSItemsCrosstab		X		
10.1	Create macro that will open query Age25+			X	
10.2	Create macro that will open query Bfn25Sprints			X	
10.3	Create macro that will open query DHS Items Crosstab			X	
10.4	Create macro that will open query DHS Items			X	
10.5	Close Switchboard = Macro not built on previous object	X			
11.1	Create Switchboard	X			
11.2	Switchboard item built on query and macro: Open Age25+				X
11.3	Switchboard item built on query and macro: Open Bfn25Sprints				X
11.4	Switchboard item built on query and macro: Open DHS Items Crosstab				X
11.5	Switch to 2nd switchboard = Macro not built on previous object	X			
11.6	Close Switchboard = built on macro		X		
11.7	Open AthletePerf form = not built on previous objects	X			
11.8	Open ItemSteward form = not built on previous objects	X			
11.9	Open AthleteItem form = not built on previous objects	X			
11.10	Switch back to queries	X			
11.11	Ampersands	X			
12.1	AthletePerf form not resizable	X			
12.2	Hide Minimise/Close buttons in AthletePerf form	X			
12.3	Deactivate Close button	X			
12.4	Deactivate Record selectors	X			
12.5	Deactivate Navigation buttons	X			
13	Create Close button to close AthletePerf form	X			

3.7   CONCLUSION

This chapter focused on Bloom's taxonomy levels. Due to the practical nature of the subject INL20DB, test items are marked as being on the application, synthesis or analysis level.

Keywords that can be used by the assessor to draw up test items were tabled. The curriculum modules, with the assessment outcomes linked to the subject areas/object types of Microsoft Access, were discussed. Finally a table was created to help the assessor to determine the level of a test item.

In Chapter 4, existing test item-generating software packages are evaluated.

3.8   REFERENCES

[1]   Bloom, B.S. & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals, by a committee of college and university assessors.* Handbook 1: Cognitive Domain. New York: Longman.

[2]   Anderson, L.W. & Krathwohl, D.R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman.

[3]   Hay, H.R. (Ed.); Mahlomaholo, M.G.; Van der Merwe, B.C.; Brussouw, S.M.; Lampbrecht, G. & Badenhorst, J.J.C. (2004). *Manual for Teaching and Learning*. 1st Edition. Bloemfontein: Central University of Technology, Free State, Unit for Academic Development.

[4]   Bast, K.; Cygman, L.; Flynn, G. & Tidwell, R. (2006). *Succeeding in business with Microsoft Office Access 2003: A problem-solving approach.* Canada: Thomson Course Technology.

[5]   Bocij, P. & Greasley, A. (1999). *Can computer-based testing achieve quality and efficiency in assessment?* International Journal of Educational Technology, 1(1): 1-18.

[6]   Lubbe, M. & Benson, S. (2005). *Let's Do It! ICDL the Practical Way: Microsoft XP*. Meyerton: Let's Do It!

[7]   Wikipedia (2008). *Curriculum*. Available online at: http://en.wikipedia.org/wiki/Curriculum (Accessed: 2008/03/14)

[8]   MacGillivray, H. (2005). *Criteria, standards and assessment in statistical education*. Brisbane: Queensland University of Technology, School of Mathematical Sciences.

CHAPTER 4 : AVAILABLE SOFTWARE PACKAGES

4.1   INTRODUCTION

Students at the CUT are assessed to determine whether they have mastered all the skills specified in the curriculum outcomes. The assessor uses test items to assess the students. A test-distribution model has been suggested to guide the assessor during the assessment process.

Chapter 3 discussed the key categories from Bloom's taxonomy and drew links between these categories and the curriculum outcomes for second-year IT students in the practical component of the subject Information Systems II.

Each curriculum outcome module was linked to a subject area/object type. Keywords that the lecturer can use to draw up test items were grouped alongside Bloom's application, analysis and synthesis levels. These keywords are also found in the curriculum outcomes of each module.

Multi-level test items were discussed, as were means of determining the level of a test item. In the assessment example that was given to students (Appendix A:A1), four levels were identified.

The assessor must have a software product to guide him/her when drawing up test items. A helpful product will link a test item to a specific curriculum outcome, a subject area/object type, a key verb, a difficulty level, as well as a cognitive level. This will enable useful single- and multi-level test items to be developed.

Keeping this information in mind, it is now possible to start evaluating different software packages. Test-building software will be evaluated to determine whether the above-mentioned attributes are available. The software will also be evaluated for other attributes, as mentioned in Chapter 2.

4.2　CRITERIA

In order to evaluate existing packages, the criteria identified in chapter 2 are applied, namely:

4.2.1　Formatting

Potgieter [1] pointed out the need for test-building software that will allow the assessor to format test papers with ease. Automatic numbering and the ability to export a test to a well-known program like Microsoft Word will be an advantage. Spellchecker, as well as Find and Replace options, are important characteristics of such a program (Table 4.1).

Table 4.1: Criteria used to evaluate software packages that can create test items

Formatting	Type of Test Item	Difficulty Level
Presentation Method	Subject Area / Object Type	Software Functionality
Randomisation	Curriculum Outcomes	Bloom's Taxonomy Levels
Help	Multi-Level Test Items	

4.2.2　Presentation method

A test must be presented in such a way that it will be possible to use it as a paper-based test and as a computer-based test [1].

4.2.3 Randomisation

Randomisation is an important characteristic as far as security is concerned. When students receive the same test, the randomisation of test items can minimise cheating. If different sets of tests can be drawn from one databank and the test items can be randomised in each set, then it becomes difficult for students to become familiar with a test paper ahead of time. This results in databank test items with high integrity [1].

4.2.4 Help

The Help feature will guide the lecturer in compiling tests [1].

4.2.5 Types of test items

There is a need for test items other than the normal True/False or the typical Multiple-Choice test items. Packages will be evaluated to determine the variation in the types of test items in each package [1].

4.2.6 Subject area / Object type

If an outcome stipulates that the student must be able to save an object, e.g. a table, then such an item outcome can be marked as "Save a Table". The subject area/object type is closely linked to the curriculum outcomes and is one of the main objects of a database, namely the database itself, a table, query, form, report, page, macro and/or module. Test items that are grouped according to a subject area/object type can help the lecturer to draw up a paper with the same number of test items from each area. An analysis of a student's test can guide him/her in areas not yet mastered. Linking test items to the curriculum outcomes and the subject area/object type can assist with this type of analysis (Chapter 2, paragraph 2.6).

4.2.7 Curriculum outcomes

Test items that are categorised according to curriculum outcomes can assist the lecturer in drawing up an equally representative paper where all the curriculum outcomes are evaluated (Chapter 2, paragraph 2.6).

4.2.8 Multi-level test items

Multi-level test items allow the assessor to create test items that build on one another. Randomisation cannot simply be applied to any multi-level test item. A level-2 test item must follow a level-1 test item and so on. Table 4.2 displays level-1 and level-2 test items. Test items 1.1 and 1.2 can only be randomised in the group of test items linked to test item 1. Test items 2.1 and 2.2 can also only be randomised in the group of test items linked to test item 2 (Chapter 2, paragraph 2.6).

Table 4.2: Randomisation of multi-level test items

Level 1	Level 2	Randomised
1. Create table Client	1.1) Add field ClientNum	Yes
	1.2) Add field ClientName	Yes
2. Create table Product	2.1) Add field ProductNum	Yes
	2.2) Add field ProductDesc	Yes

4.2.9 Difficulty level

The lecturer will find it valuable to mark the difficulty level of test items, as it will enable him/her to maintain a specific standard in each paper. It is challenging to determine the difficulty level of a test item in advance, and so it is better to rate the difficulty level of each test item after the students have written the test. The difficulty level can then be updated in the databank (Chapter 2, paragraph 2.5.1).

4.2.10 Software functionality

Software functionality tests the suitability of the test-developer software product (Chapter 2, paragraph 2.5.1.3.2). The amount of resources it uses is evaluated, along with the software's ease of use and the trustworthiness of the results. For purposes of this research, the results refer to the ability of the software to generate a paper with all the attributes selected by the lecturer.

4.2.11 Bloom's taxonomy levels

According to the revised taxonomy of Bloom, a second-year exam paper must consist of 30% remembering skills, 20% understanding, 20% application, 10% analysis, 10% evaluating and 10% creating skills [2]. If each test item in the databank is linked to a category in Bloom's taxonomy list, then it will be possible to control the number of test items from each category for a specific year group (Chapter 3).

4.3   AVAILABLE PACKAGES

There are a number of test-developer packages available that allow the lecturer to draw up tests and publish them for students to complete. The focus in this chapter is on specific test-developer software packages, namely ExamView Pro [4], Respondus 3.5 [5], WebQuizXP [6], Articulate Quizmaker 2.0 [7], Question Mark Designer [8], Adaptex [9], Random Test Generator PRO [10], Interactive Question Server [11], Adit Test Desk [12], TCExam [13], Easy Test Maker [14], Give Testv2 [15] and EasyCast [16].

The above-mentioned packages were evaluated according to the characteristics mentioned in Table 2.5, Chapter 2. These tables exclude test-item types like True/False. Test-item types are discussed individually under each package.

Table 4.3: Evaluation of test-delivery software packages 1-7

Category	Description	Package						
		ExamView Pro	Respondus	WebQuiz XP	Articulate Quizmaker	Question Mark Designer	Adaptex	Random Test Generator Pro
Formatting	Numbering	1	1	1	1	1	1	1
	Find/Replace	1	1	1	1	1	1	1
	Spellchecker	1	1	1	1	1	1	1
	Paper Compilation	1	1	1	1	1	1	1
	Exporting paper for editing	1	1	1	1	1	1	1
Randomisation	Arranging items	1	1	1	1	1	1	1
Help	Giving guidance in drawing up a paper	1	1	1	1	1	1	1
Difficulty Level	Determining the difficulty level of a test item	1	0	0	0	0	1	0
Bloom	Category	1	0	0	0	0	1	0
Curriculum Outcomes	Matching test item with outcomes in a curriculum	1	0	0	0	1	1	0
Subject Area/ Object Type	Linking items to subject area/object type	1	0	0	0	1	1	0
Multi-Level Test Items	Level 1	0	0	0	0	0	0	0
	Level 2	0	0	0	0	0	0	0
	Level 3	0	0	0	0	0	0	0
	Level 4	0	0	0	0	0	0	0
Presentation Method	CBA test available on network	1	1	1	1	1	1	1
	Option to print test	1	1	1	1	1	1	1

Software Functionality	Is software working?	1	1	1	1	1	1	1
	How rapidly does it generate tests and results?	1	1	1	1	1	1	1
	What is the amount of resources used?	1	1	1	1	1	1	1
	Is it easy to install?	1	1	1	1	1	1	1
	Is the software easy to use?	1	1	1	1	1	1	1
0 = characteristic not available 1 = characteristic available								

Table 4.4: Evaluation of test-delivery software packages 8-13

Category	Description	Package					
		IQS 1.5	Adit Testdesk	TCExam	Easy Test Maker	GiveTestv2	EasyCast
Formatting	Numbering	1	1	1	1	1	1
	Find/Replace	1	1	1	1	1	1
	Spellchecker	1	1	1	1	1	1
	Paper compilation	1	1	1	1	1	1
	Exporting paper for editing	1	1	1	1	1	0
Randomisation	Arranging items	1	1	0	1	1	1
Help	Giving guidance in drawing up a paper	1	1	1	1	1	1
Difficulty Level	Determining the difficulty level of a test item	0	1	1	0	0	1

Bloom	Category	0	0	0	0	0	1
Curriculum Outcomes	Matching test item with outcomes in a curriculum	0	1	1	0	1	1
Subject Area/ Object Type	Linking items to subject area/object type	0	0	1	0	1	1
Multi-Level Test Items	Level 1	0	0	0	0	0	0
	Level 2	0	0	0	0	0	0
	Level 3	0	0	0	0	0	0
	Level 4	0	0	0	0	0	0
Presentation Method	CBA test available on network	1	1	1	1	1	1
	Option to print test	1	1	1	1	1	1
Software Functionality	Is software working?	1	1	1	1	1	1
	How rapidly does it generate tests and results?	1	1	1	1	1	1
	What is the amount of resources used?	1	1	1	1	1	1
	Is it easy to install?	1	1	1	1	1	1
	Is the software easy to use?	1	1	1	1	0	0
0 = characteristic not available 1 = characteristic available							

4.4   ExamView PRO

4.4.1 Background and characteristics

The first package under discussion is a product from ExamView Pro. A suite of packages is available that will help the user to administer tests and test results. With ExamView Test Builder [4] the assessor can draw up tests, while ExamView Test Generator groups test items according to the test-item type. ExamView Test Manager reads online test results, generates reports and includes password protection for test access. A student can also get customised feedback, based on his/her responses. A test can be printed or published using ExamView Test Player software.

Test items can be formatted in a full-featured word processor, using various predefined styles. The formatted test items can include pictures, graphs, charts, tables, and sound and movie files. Multiple versions of the same test can be generated, because randomisation is available. A test can have a mix of test-item types. An item can be grouped according to test-item type, curriculum outcome, subject area/object type, as well as learning objective. There is a miscellaneous area where test items can be grouped according to other features.

Plain-paper scanning is supported where bubble optical answer sheets can be printed, allowing scanners to mark these tests. A high-contrast and easily recognisable shape like a circle is used. A familiar way to answer these tests is by using an HB2 pencil to colour in a circle, after which a scanning machine grades the sheet. Customised feedback can also be given to a student based on his/her responses. Cartesian, Polar and Number-Line graphs can be inserted, which makes ExamView Pro ideal for mathematical questionnaires.

4.4.2 Types of test items available

ExamView Pro supports 14 types of test items:

- True/False: A statement is given that can be either true or false.

- Modified True/False: Either True or False can be selected, but an explanation must be given as to the reasons why that specific option was selected.
- Multiple test items:
  - Multiple-Choice: Only one of several choices is correct.
  - Multiple-Response: More than one option may be selected.
  - Bimodal: Only one of several choices is correct, and a detailed explanation must be given of the reasons why that specific option was selected.
  - Numeric Response: The answer will be a numeric typed-in value.
- Numeric Multiple-Choice test items using:
  - Cartesion Graph: Used to determine each point uniquely in a plane, through two numbers usually called the x-coordinate and the y-coordinate.
  - Number-Line Graph: The line graph is another graph that represents the adjacencies between edges of the original graph.
  - Polar Graph: A two-dimensional coordinate system in which each point on a plane is determined by an angle and a distance.
- Completion: To complete a sentence by filling in one or more words.
- Matching: There are two columns of choices, and each item in the first column must match a choice in the second column.
- Open-ended test items:
  - Short-Answer: To complete a statement by filling in a single word.
  - Essay: To complete a statement by filling in a description or writing a paragraph.
  - Problem: The student is given a case study.
- Dynamic: Answering choices and values that are automatically substituted to deliver dynamic instead of static test items.

These test items can be saved in test-item banks and can be updated over the internet.

4.4.3 Shortcomings

It is not possible to link test items to Bloom's taxonomy levels. The product does not allow the lecturer to draw up multi-level test items that can be randomised.

4.5   Respondus

4.5.1 Background and characteristics

The second package under discussion is Respondus [5], which is used to create test items and to apply formatting and settings. Test items can be published to a program called WebCT, which allows the lecturer to publish tests online from where students can write the tests and the tests can be assessed.

Test items can be created within the package or be imported as plain text or in rich-text format from another Word-based document. Test items can be formatted within Respondus together with a point value linked to every test item. Feedback to students can be enabled.

A handy feature is test-item sets. A test-item set can contain any number of test items within the set that can be randomised. For example, a lecturer can create a test-item set containing eight test items from which only four test items will be selected. These four test items can then be randomised.

There are extra features available like the archiving of files and folders, and item visibility to students. The test-item delivery method determines whether test items are delivered one at a time, and this can be revisited by students.   The duration of the paper can be set, as can the number of attempts by a student to write a test. The date and time of the availability of the test can be set together with an event that will automatically be created in a calendar on WebCT. Security is also

addressed. Other settings like score releases and the way in which the results are released to students are available. Summary statistics of students can be retrieved from WebCT, including the summary statistics and answer distribution for each test item. It is a very well-developed software package with many usable features.

4.5.2 Types of test items available

Respondus supports 9 types of test items:

- True/False: A statement is given that is either true or false.
- Multiple-Choice: Only one of several choices is correct.
- Multiple-Response: More than one option may be selected.
- Calculated: Calculations must be done.
- Short Answer: To complete a statement by filling in a single word.
- Paragraph: To complete a test item by writing more than one word.
- Fill-in-the-Blank: To complete a sentence by filling in a single word.
- Jumbled Sentences: To arrange sentences in the correct sequence.
- Matching: There are two columns of choices, and each item in the first column must match a choice in the second column.

4.5.3 Shortcomings

Although it is possible to create sets of test items, the product does not visually link one set to another. Test-item settings do not include a link with Bloom's knowledge levels, the curriculum outcomes, or the subject area/object type.

4.6 WebQuiz XP

4.6.1 Background and characteristics

WebQuiz XP allows the lecturer to create assessments that can be published [6]. The assessor can add, edit and delete test items online. Each test item can have a different score according to the answer given by the student. Test items can be categorised into topics and be randomised. It is possible to select a certain number of test items from each topic. Test items can be displayed one per page or all together on one page. Pictures, graphs and equations can be added. The lecturer can preview tests in an online environment. Existing tests can be imported as .txt files (text files without formatting) and used in online tests. Test items can also be exported as .rtf (rich-text format) files, which will allow the lecturer to hand out the test printed on paper.

The graphical layout of tests can be changed by using an available template. Lecturers will have access to the source code, which will allow them to apply changes according to their needs. It is possible to add, edit and delete users and to determine the number of log-ins per test.

Results are saved in Microsoft Access and can be exported to Microsoft Word or Microsoft Excel or be turned into simple text files. Lower and upper bounds can be added to tests, with messages to students informing them whether or not they have passed the test in question, based on the upper- and lower-bound messages. The student can be sent a custom e-mail message with his/her test results. A memorandum can automatically be forwarded to a student after completion of a test. A certificate can be generated for a student the moment his/her score exceeds the predefined passing value.

4.6.2 Types of test items available

WebQuiz XP supports five types of test items:

- Multiple-Choice: Only one choice of several is correct.
- Multiple-Answer: More than one option may be selected.

- True/False: Only one choice out of several is correct.
- Fill-in-the-Blank: One or more words, whole numbers, or numbers of mixed value may be inserted.
- Essay: To complete a statement by writing a description or paragraph.

4.6.3 Shortcomings

It is not possible to draw up multi-level test items. Test-item settings do not include attributes like Bloom's knowledge levels, the curriculum outcomes, the subject area/object type, or the difficulty level.

4.7   <u>Articulate Quizmaker 2.0</u>

4.7.1 Background and characteristics

Articulate Quizmaker 2.0 [7] is a quiz creator with many usable features. The lecturer can incorporate images and flash movies in test items. Extra features include the option to customise buttons and labels to the lecturer's choice of text, colours, sound and language, or to brand tests with custom colours and sounds. A time limit can be set for the completion of the test. Test items can be published to a Word document for a printable version, or be published to PowerPoint.

Students can navigate through test items in any order, which enables them to skip test items for later review. They can finish a questionnaire without answering all the test items. The lecturer can choose the amount of information to be displayed for each test item, such as the maximum points per test item and the points awarded to the student for his/her answer. Feedback, based on the student's responses, can be selected to guide the student in answering test items. An alert message will appear if a student fails to select an answer before submitting the question. Extra information on the test item can also be given for guidance purposes. Test items can be submitted one by one or all at once after

completion of the questionnaire. Results can be displayed, as can actions that the student can follow based on his/her score. A student's results can be forwarded to him/her in an e-mail, and the student is allowed to print a record of his/her test results.

4.7.2 Types of test items available

Articulate Quizmaker 2.0 supports 21 types of test items:

- True/False: A statement is given that can be either true or false.
- Multiple-Choice: Only one choice of several is correct.
- Match
    - Matching: Drag and drop items to match item on left with item on right.
    - Matching Drag and Drop: Drag and drop items to arrange in sequence.
- Hotspot: Click in a specific area to indicate your choice.
- Likert Scale: Choose a response that best represents your opinion.
- Word Bank: Drag and drop the correct word response into an empty text box.
- Fill-in-the-Blank: Enter one or more words.
- Which Word: Drag and drop the word that best represents your opinion.
- Short Answer: Enter a short comment or opinion.
- Essay: Enter an essay response.
- Explanation: View reference information, which can include text, images, links and attachments.
- Pick:
    - Pick One: Choose a single item from multiple options.
    - Pick Many: Choose several items from multiple options.

- Rank:
  - Ranking Drop-down: Select items from drop-down menus to rank in preferential order.
  - Ranking Drag and Drop: Drag and drop items to rank in preferential order.
- Numeric: The answer is a numeric typed-in value.
- How Many: Enter a numeric response.

4.7.3 Shortcomings

Articulate Quizmaker 2.0 does not allow the lecturer to draw up multi-level test items.   Test items do not include Bloom's knowledge levels, the curriculum outcomes, or the subject area/object type. Test items are not linked to a difficulty level.

4.8 <u>Questionmark Perception</u>

4.8.1 Background and characteristics

Questionmark Perception [8] enables the assessor to create, modify and delete tests, assessments and surveys directly from the Perception server if he/she does not wish to install the software on his/her computer. Test items can be organised by topic and can be randomised. These items can be exported to Microsoft Word for editing. The test items are then mailed to an administrator who publishes them in Questionmark Perception.

Twenty built-in templates are available that will allow the lecturer to change the format of tests according to his/her needs. These templates can be modified in Notepad or Template Editor.   Equations, graphics, multimedia files, flash, sound, videos and URLs can be used in test items. More than one author can

save test items in the databank, and if different authors are working on items in the same area, naming conventions will be used in order to keep a particular author's work separate from that of the others.

Test items can be adapted based on the students' answers, which makes this an adaptive model. Time limitations can be set on the assessment, and feedback can be given to the students at the end of the test. A general message can be generated to give the student feedback on any errors made, while a specific message can be created for each error in a multiple-choice test.

Reports can be generated that include the answers and scores of an individual student in a single test. Grade-book reports look at the results of students over the course of a number of tests. Item-statistic reports provide statistics based on the responses to individual test items, which make it possible to look at test-item validity and efficiency. Test-item validity tests the quality and efficiency of a test item, while test-item efficiency tests whether an item tests what it is supposed to test.

4.8.2 Types of test items available

QuestionMark Perception supports 22 types of test items:

- Drag-and-Drop: Drag and drop items into the correct position.
- Essay Test Items: Answer in paragraph format of up to 30 000 characters.
- Explanation Screens: View a text or graphic prior to answering a series of test items.
- File Upload: Complete an assignment and then upload it.
- Fill-in-the-Blank: Fill in one or more missing words in a statement.
- Hotspot: Make a selection by clicking on certain pictures or areas of a picture.

- Knowledge Matrix: Select one answer for each of several multiple-choice statements.
- Survey Matrix: From multiple rows of Likert-scale test items in column format, choose the response that best represents your opinion.
- Likert Scale: Choose a response that best represents your opinion.
- Matching: There are two columns of choices, and each item in the first column must match a choice in the second column.
- Multiple-Choice: Only one of up to 40 options is correct.
- Multiple-Response: More than one choice can be correct.
- Numeric Test Items: Respond with a numeric value or a value within a range.
- Pull-down List: Use a pull-down list to match a series of statements with one another.
- Ranking: Rank choices numerically.
- Select-a-Blank: Use a pull-down list to select an answer to fill in a blank in a statement.
- True/False: The statement given can be either true or false.
- Word Response: Enter one or more words that match the answer.
- Yes/No: Only one option is correct.
- Adobe Flash: Adobe Flash is used to create a test item, and when the student answers that test item, the result is recorded within the answer database.
- Adobe Captivate Simulations: A simulation is created and the student is scored on multiple interactions.
- Spoken Response: The student's spoken responses can be recorded and processed along with other test items.

4.8.3 Shortcomings

Questionmark Perception does not allow the assessor to draw up multi-level test items. Test items do not include Bloom's knowledge levels and do not allow the lecturer to set the difficulty level of a test item.

4.9   Adaptex

4.9.1 Background and characteristics

Adaptex [9] consists of four items, namely an item-banking module, a test-specification module, a test-administration module (both server-based and web-based) and a test-results module. Adaptex uses Microsoft Access to store item-bank information, test specifications and examinee results. Adaptex was developed in such a way that other software that is likely to be on any computer can also be used in the development of test items. The assumption is that most computers have Microsoft Access software installed, which allows the lecturer to format test items in Microsoft Word.

Item difficulty, item discrimination, content categories and the cognitive level of test items are attributes that can be added to each test item. Item feedback can be provided along with appropriate messages to students who have passed or failed a test. A time limit can be set for the test. Because the test developer can store a variety of information about each item, the student will be allowed to respond to test items by typing in only a stem instead of the full or derived word. Cut-off scores can be applied that will terminate a test session once a student has scored a specified number of points. Passwords can be set to restrict access to tests.

Three different scoring methods can be used, namely: Number of Items Correct, Item Response Theory (IRT) maximum likelihood estimate, or Percentage of Correct Answers. In an IRT-based test, the next test item that the student will

receive will be based on his/her responses to previous items. It is possible to administer test items in a fixed or randomised order.

The test-results module can include the student's name, ID number, the date and time of the test, his/her response to each test item, whether each response is correct or incorrect, the final score, and the time taken by the student to respond to each item.

4.9.2 Types of test items available

Adaptex supports 1 question type:

- Multiple-choice items can be developed, which can include multimedia clips like audio and video, graphics and reading passages.

4.9.3 Shortcomings

The product does not allow the lecturer to draw up multi-level test items.

4.10   Random Test Generator PRO

4.10.1 Background and characteristics

Random Test Generator PRO version 8.2 [10] allows the lecturer to develop test banks of test items. There is no limitation to the number of test banks or test items per test bank. Test items can be grouped by topic, and different tests can be generated for students containing different test items on the same topic or the same test items that have been randomised.

Random Test Generator PRO uses an interface similar to Microsoft Word, which makes the software easy to use because it is familiar to many users. Test items,

student tests and item analysis can be saved as rich-text-format or text files, which makes them accessible to Microsoft Word and other word-processing software. This means that editing can take place outside Random Test Generator PRO. The software allows the use of graphics, animation, movies and sound in any test item. Test items may be read out aloud by the software or be read from the screen.

Databanks are maintained in Microsoft Access for reliability and security purposes. Log-in passwords for students and lecturers can be created. Test administration allows the lecturer to set the time limit within which a paper must be completed. Tests can be printed and students can submit tests via e-mail for assessment. Track is kept of every test item answered by a student. These answers can be used to test the validity of items.

Standard versus practice tests can be scheduled. With standard tests students will receive feedback from the lecturer when the feedback option has been selected by the lecturer. With practice tests, students will always receive feedback. During practice tests, the students will be guided with online feedback when answering test items.

4.10.2 Types of test items available

Random Test Generator PRO supports 5 types of test items:

- Multiple-Choice:
    o Single Answer: Only one option is correct.
    o Multiple-Answer: More than one option might be correct.
    o True/False: A statement given is either true or false.
    o Fill-in: One or more words must be entered in the blank space.
- Essay Type: The answer can be a short sentence or paragraph.

4.10.3 Shortcomings

The product does not allow the lecturer to draw up multi-level test items. Test items do not include information concerning the different knowledge levels. Items cannot be linked to curriculum outcomes.

4.11 <u>Interactive Question Server (IQS) 1.5</u>

4.11.1 Background and characteristics

WebMCQ has a product available known as IQS 1.5 [11], which makes it easy to create and manage online tests. Test items can be imported, exported, edited and copied. Multimedia and graphics can be incorporated. Fifty sets of test items can be set up by the lecturer. It is possible to switch between sets of test items at any time, which enables the assessor to allow one student group to write a particular set of test items while another group writes a different set. Workgroups can be created that allow students to create their own test items. As an alternative, WebMCQ is willing to reskin the application according to the individual's personal needs. WebMCQ has released an updated version known as Ensignia, which has more item templates, randomisation abilities, and the option to facilitate the online submission and marking of essays by human assessors.

Feedback after answering a test item can be withheld until the student wishes to see it. Time limits can be applied within which the students must complete a test. An analysis of a student's performance can be conducted and presented to him/her.

Statistics are available on how many people have accessed test items, how often and when. Individual statistics like the final score, section scores and individual test-item responses are available. Reports can be filtered by date and group. Diagnostic feedback can be provided, which will identify the student's strengths and weaknesses in terms of knowledge.

Logs of all activities make it possible to identify whether any attempts were made to improperly access test items. Built-in security determines who may have access to tests, for how long and when.

4.11.2 Types of test items available

Interactive Question Server (IQS) 1.5 supports 7 types of test items:

- Multiple-Choice: Only one answer amongst many is correct.
- True/False: A statement that is given is either true or false.
- Multiple Correct-Response: More than one answer amongst many can be correct.
- Likert Scale: Choose a response that best represents your opinion.
- Short Answer: Fill in two or more words, but not a paragraph.
- Fill-in-the-Blank: Fill in one or more missing words in the statement.
- Hotspot: Make a selection by clicking on certain pictures or areas of a picture.
- Drag-and-Drop: Images are dragged and dropped into the correct position.

4.11.3 Shortcomings

The product does not allow the lecturer to draw up multi-level test items with links between test items on the different levels. The knowledge level of a test item

cannot be selected, nor can the curriculum outcomes or the subject area/object type. It is not possible to identify the difficulty level of a test item.

4.12   **Adit Testdesk**

4.12.1 Background and characteristics

Adit Testdesk [12] has several programs available in its suite, namely:
- Scriptmaker, which allows one to create script tests.
- Editor, which can be used to create and edit tests and give test permission.
- Testserver, which manages users.
- Tester, which runs a test with a handy interface and allows the user to go to the next or previous test item, to view audio and video, to add graphics, to add tips to a test item, and to bookmark a test item in order to return to the item if need be.
- Testclient, which allows tests to run on a LAN.
- Reporter, which allows the assessor to apply filtering to process the exact test results needed and to print and export results.

Test items can be imported or exported, and a built-in text editor is available. The assessor can use the same pool of test items with variations in different tests by the use of profiles. Test items can be sorted by topic and randomised when presented to the student. The order of test items can be adjusted depending on the student's answers. This is an adaptive test module, although the order of test items can also be predefined by the assessor. Each test item can have its own weight, and conversion of item weights after the test has been written is also possible. Test items with multiple correct answers use logical expressions to confirm the validity of an answer. Messages can be activated for the student

when answering a test item. A time limit can be set within which a student must complete a test or test item, and it is also possible to track test versions. Tests can be printed and handed out to students. Moreover, a test can be protected from unauthorised access by encrypting test files and students' results and by using passwords with different levels of access rights.

The appearance of the test can be customised by using text, images, tables and Object Linking and Embedding (OLE) objects. Different fonts, styles, dragging, resizing and transparency effects are available, and the language of the interface can also be changed.

Various templates are available for test reports. The assessor can also analyse test results based on the different topics, which will allow him/her to analyse the difficulty level of a test item.

4.12.2 Types of test items available

Adit Testdesk supports 14 types of test items:

- True/False: A statement that is given can be either true or false.
- Multiple:
  - Multiple-Choice: Only one choice out of several is correct.
  - Multiple-Response: More than one option can be selected.
- Matching Test Items: There are two columns of choices, and each item in the first column must match a choice in the second column.
- Sequence: A list of choices must be placed in the correct order.
- Fill-in-the-Blank: An answer must be typed in to be assessed by the use of expressions.
- Numeric:
  - Integer Numeric: The answer will be a numeric typed-in value.

- - Float Numeric: The answer will be a numeric value with a decimal fraction.
  - Hotspot: The student must click in a specific area to indicate his/her choice.
  - Draw:
    - Draw Point: A specific coordinate must be selected as the indicated choice.
    - Draw Circle: A circle must be drawn where the coordinates of the radius and the centre of the circle will determine the correctness of the answer.
    - Draw Segment: A rectangle must be drawn where the coordinates of the end points of the lines connecting to the diagonal angles determine the correctness of the answer.
    - Draw Rectangle: A rectangular segment must be drawn.

4.12.3 Shortcomings

The product does not allow the lecturer to draw up multi-level test items. Test-item settings do not include knowledge levels.

4.13 TCExam

4.13.1 Background and characteristics

TCExam [13] is free, open-source Web-based CBA software that allows the assessor to schedule and deliver tests and create reports. An open-source feature allows any person to access code and resolve problems, or to request help from any vendor. The software is freely available, which is probably why it is a more frequently used product in developing countries. The software has been translated into nine different languages.

The two main sections in the software are the public and the administration areas. The user will use the public area to log in with a username and password in order to access the tests. The user may move between test items in any order, return to test items to change answers, and terminate a test at any time.

The administration area contains the interfaces that allow the lecturer to manage the system. From here the lecturer will manage the users and the database and also generate tests and results. A time limit can be set on a test. User groups, test items and results can be directly exported or imported in CSV (Comma-separated Values) or XML (eXtensible Markup Language) format, or as a PDF (Portable Document Format) file.

Test items can be grouped into an unlimited number of topics, and a topic can contain an unlimited number of test items. Each test item can have an unlimited number of alternative answers with a specific difficulty level. A random number of test items with certain characteristics can be extracted from each topic or group of topics. Text formatting, multimedia objects like audio and video, and mathematical formulas are supported. A test can be printed. Test activities can be monitored in real time and be changed while the student is busy writing the test. A test can be halted and restarted, or the duration of a test can be extended. Grading of tests happens in real time and the students' results can be revealed to them directly on completion of the test. The results can be forwarded to a student by e-mail, and test statistics can be exported, printed and/or saved.

4.13.2 Types of test items available

TCExam supports four types of test items:

- Multiple-Choice Single-Answer: Only one choice of several is correct.

- Multiple-Choice Multiple-Answer: More than one choice of several can be correct.

- Ordering Answer: Answers must be placed in a specific sequence.

- Text Answer: The answer can be a word, phrase, sentence, paragraph or essay. Longer test items like long paragraphs or essays are graded manually.

4.13.3 Shortcomings

The product does not allow the lecturer to draw up multi-level test items. Test items are not linked to Bloom's taxonomy levels, and the randomisation of test items is not possible.

4.14 <u>Easy Test Maker (ETM)</u>

4.14.1 Background and characteristics

Easy Test Maker (ETM) is a free online test generator [14]. Limited formatting can be applied within ETM, although a test can be downloaded as a Word document where advanced formatting features are available. Instructions can be included for each test item. It is possible to divide a test into multiple sections. For each test, two alternate versions can automatically be created. The test items, as well as the possible answers, can be randomised. Test items from different sections will not be mixed. If the assessor wishes to create a new test, he/she can copy test items from existing tests and paste them into the new test, or directly add new test items to the new test. A test can be printed along with an automatically generated answer sheet.

4.14.2 Types of test items available

Easy Test Maker supports five types of test items:

- True/False: A statement is given that can be either true or false.

- Multiple-Choice: Only one option out of several is correct.

- Fill-in-the-Blank: The answer can be one or more words to be filled into a blank space.

- Matching: There are two columns of choices, and each item in the first column must match a choice in the second column.

- Short Answers: The answer can be a short sentence or paragraph.

4.14.3 Shortcomings

Easy Test Maker does not allow the assessor to link a difficulty level to a test item. It is not possible to draw up multi-level test items. The knowledge level of a test item cannot be selected, nor can the curriculum outcomes or the subject area/object type.

4.15   GiveTestv2

4.15.1 Background and characteristics

GiveTestv2 [15] allows the lecturer to create, administer and analyse tests. Formatting can be applied to test items in almost any text document, after which test items can be imported. Formatting like graphics, formulas, animation, video and audio is also available. Test items and their answers can also be randomised. The assessor can determine a grading scale and password to get access to the test items. The test items are saved in test banks. Administrators, instructors, operators and user groups can be created where each group has a different set of permissions. A test is assigned to a specific group, and automated feedback can be provided based on each student's response. The curriculum outcome of a test item can be linked to each test item. Test results can be forwarded to users via e-mail, and statistics can be gathered from the

students' answers to each test item. Reports and certificates can be generated. Web hosting for tests is provided on the servers of the GiveTestv2 developers.

4.15.2 Types of test items available

GiveTestv2 supports 5 types of test items:

- True/False: A statement is given that can be either true or false.
- Multiple-Choice Single-Answer: Only one choice out of several is correct.
- Multiple-Choice Multiple-Answer: More than one option can be correct.
- Short Answer: The answer can be a word or short sentence.
- Essay: The answer can be a paragraph.

4.15.3 Shortcomings

GiveTestv2 does not allow the assessor to link a test item to a difficulty level. It is not possible to draw up multi-level test items, and the knowledge level of a test item cannot be linked to that test item.

4.16   EasyCast (WebX)

4.16.1 Background and characteristics

EasyCast [16] is an application developed to test students in multiple subjects. A detailed walk-through tutorial with video is provided to help the assessor in using the software. Text- and graphic-based test items are supported. An assessment is divided into different sections into which the assessor can place the test items. Test items on the same level will be placed in the same section. Once the students have completed one section they must move on to the next section. The test items can be randomised.

The students' responses to different test items and to each objective can be analysed. The students' results can be exported to Microsoft Excel for further analysis. Different types of reports can be generated, such as class lists, student performance reports and student item reports.

4.16.2 Types of test items available

EasyCast (WebX) supports 4 types of test items:

- Multiple-Choice Single-Answer: Only one choice out of several is correct.
- Multiple-Choice Multiple-Answer: More than one choice can be correct.
- Image-Based Multiple-Choice: Images form part of a test item where more than one choice can be correct.
- Virtual Environment Test Items: Students work in an e-mail, word-processing, spreadsheet, database and Windows management environment.

4.16.3 Shortcomings

There is no link between the multi-level test items.

4.17   CONCLUSION

Chapter 4 discussed different software packages that allow an assessor to draw up test items. It became clear that no software package will meet all the needs of INL20DB lecturers teaching Microsoft Access. Each tool forms a useful part of the collective strategy used by assessors. It will be possible for a lecturer to identify the package that most closely meets his/her needs. It is therefore not appropriate to rate packages as either "good" or "bad".

It is important to be able to create single- and multi-level test items when assessing databases in the subject INL20DB. A shortcoming in all the packages discussed (Table 4.3 and Table 4.4) is the inability to create multi-level test items and to link single- and multi-level test items. To a lesser degree there is a lack of ability to link a difficulty and cognitive level and a curriculum outcome and subject area/object type to a test item.

In most of the packages discussed, formatting is not an issue, because the assessor can either export the paper for formatting in a Word-based environment, or the package itself has the necessary formatting capabilities.

The software packages discussed all functioned properly when being operated. The packages were easy to install and use and the TMS did not take up a lot of resources.

Chapter 5 discusses the evaluation of the students' performance in test items based on gender, population group, difficulty level, cognitive level, object type, and single- and multi-level test items. On the cognitive level, the main focus will be on test items on the synthesis and application levels.

4.18 REFERENCES

[1] Potgieter, P.H. (2001). *Evaluering van rekenaarpakkette as hulpmiddel by die opstel van vraestelle by die Technikon Vrystaat*. Bloemfontein: Technikon Vrystaat, Departement Inligtingstegnologie.

[2] RITE Group. (2008). *Teachers Helping Teachers*. Available online at: http://www.rite.ed.qut.edu.au (Accessed: 2008/01/07)

[3] Hay, H.R. (Ed.); Mahlomaholo, M.G.; Van der Merwe, B.C.; Brussouw, S.M.; Lampbrecht, G. & Badenhorst, J.J.C. (2004). *Manual for Teaching and Learning*. 1st Edition. Bloemfontein: Central University of Technology, Free State, Unit for Academic Development.

[4] Thomson South-Western. (2004). *ExamView Pro*. Available online at: http://www.swlearning.com/examview/examview_main.html (Accessed: 2008/01/07)

[5] Respondus. (2008). *Respondus 3.5*. Available online at: http://www.respondus.com/download/respondus.shtml (Accessed: 2008/01/15)

[6] Sofotex. (2003). *WebQuiz XP*. Available online at: http://www.sofotex.com/SmartLite-WebQuiz-download_L10018.html (Accessed: 2008/01/08)

[7] Articulate. (2006). *Articulate Quizmaker 2.0*. Available online at: http://www.articulate.com/downloads/freetrial-step1.aspx (Accessed: 2008/01/24)

[8]   Questionmark. (2005). *Questionmark Perception*. Available online at:
        http://www.questionmark.com/us/index.aspx (Accessed: 2008/01/24)

[9]   ICBA Services. (2004). *Adaptex Assessment System*. Available online at:
        http://www.jmu.edu/icba/prodserv/adaptexoverview.htm
        (Accessed: 2008/02/08)

[10]   Hirtle Software. (2008). *Random Test Generator PRO*. Available online at:
        http://www.hirtlesoftware.com/p_rtgpro.htm (Accessed: 2008/02/20)

[11]   MCQ International. (2008). *Interactive Question Server*. Available online
        at: http://www.mcqi.com.au/iqs (Accessed: 2008/02/28)

[12]   Adit Software. (2008). *Adit Testdesk*. Available online at:
        http://www.aditsoftware.com (Accessed: 2008/02/28)

[13]   Tecnick. (2008). *TCExam*. Available online at:
        http://www.tecnick.com/public/code/cp_dpage.php?aiocp_dp=tcexam
        (Accessed: 2008/03/05)

[14]   EasyTestMaker. (2008). *EasyTestMaker*. Available online at:
        http://www.easytestmaker.com/default.aspx (Accessed: 2008/03/07)

[15]   Iglearn. (2007). *GiveTestv2*. Available online at: http://igivetest.com
        (Accessed: 2008/03/11)

[16]   EdTech Systems. (2008). *EasyCast (WebX)*. Available online at:
        http://www.edtechsystems.com (Accessed: 2008/03/08)

CHAPTER 5 : ANALYSIS

5.1   INTRODUCTION

The main focus of this study was the creation of a test-item distribution model for INL20DB students, along with software that will allow the assessor to draw up single- and multi-level test items. These items would have to be linked to a difficulty level, one of Bloom's cognitive levels, an object, and a curriculum outcome.

Chapter 4 discussed different test-item-generating software packages and evaluated them using a combination of three existing models (Chapter 2). The conclusion drawn was that there is a need for a software product that will allow CUT lecturers teaching databases to second-year IT students to create single- and multi-level test items in assessments. These multi-level test items must have the ability to be linked to one another. Each test item must be linked to a difficulty level, curriculum outcome, object type, and one of Bloom's taxonomy levels.

Chapter 5 reflects the statistical analysis of single- versus multi-level test items, Bloom's taxonomy levels, object types, and the difficulty level of test items.   Two confounding variables, namely gender and population group, were taken into account. The results were gathered from test items based on Microsoft Access 2002. Second-year IT students in the subject INL20DB were assessed.

In this chapter the performance of the students in test items is evaluated based on gender, population group, difficulty level, cognitive level, object type, and single- and multi-level test items. On the cognitive level, the main focus will be on test items on the synthesis and application levels.

5.2 BACKGROUND

Microsoft Access 2002 is database software that is used to manipulate large quantities of data related to a specific subject or purpose [1].

The test given to the students consisted of twelve main test items, some of which were subdivided into smaller sets of tasks. In total the students were evaluated on 58 actions that they were required to complete. The subtotal of the test was 35 marks, and the students were given 80 minutes to complete the test. The test was administered during the students' normal practical instruction period. Students received a printout test (Appendix A:A1) and completed the test on computer using Microsoft Access 2002.

The curriculum outcome of each test item was based on the current Microsoft Access curriculum for second-year students at the CUT (Chapter 3). The test was comprised of single- and multi-level test items on the application, synthesis and analysis levels [2]. The difficulty levels of the test items varied and were not scientifically predetermined. It is suggested that future research should test the difficulty level of test items in order to scientifically determine their impact on single- versus multi-level test items. Each test item could be linked to an object type like a table, query, form or macro.

5.3 STUDENTS' LEVEL OF KNOWLEDGE

The performance of full-time second-year IT students in the subject INL20DB was evaluated. These students had in their first year of study already successfully completed the subject INL10DB, where they had been introduced to the basics of Microsoft Access 2002. For the first-year subject INL10DB the students had been evaluated on ECDL, while the second-year subject INL20DB involved a more in-depth study of Microsoft Access 2002. The students were

evaluated during the month of September, after completing the curriculum (Chapter 3) for second-year students.

5.4   BACKGROUND TO THE METHODOLOGY

The aim of the research was to find solutions to certain problems [2]. In this study the goal was to develop a distribution model for test items whereby second-year IT students at the CUT can be assessed to determine whether they are ready to be promoted to the next level of study.

The correlational research type was used to collect data. Correlational research attempts to find a correlation between two or more variables of the same group, for example where a correlation is found between the level of a test item and the object that was evaluated [2].

The confounding variables taken into account were population group and gender. The education system in South Africa is embedded in a Western culture with aspects that are unfamiliar to most African cultures [3]. This can have an effect on the way in which different population groups experience assessment items. In order to determine whether the results of the 60.2% males who took part in the research could distort the data, gender was selected as another confounding variable.

The questions investigated:

- Do population group and gender play a role in the correct answering of test items on the different difficulty levels, on Bloom's synthesis and application levels, on objects like tables, queries, forms and macros, and on single- versus multi-level test items?
- How do students perform in single- versus multi-level test items?

- Does object type play a role in a student's performance in single- versus multi-level test items?

- Do the application and synthesis levels of test items play a role in a student's performance in single- versus multi-level test items?

- How do students perform in test items on Bloom's application and synthesis levels?

- How do students perform in test items on different difficulty levels?

- How do students perform in test items in terms of the different objects (tables, queries, forms, macros)?

- Does an object that is linked to Bloom's application and synthesis levels have an influence on student performance?

- Does an object that is linked to the difficulty level of a test item have an influence on student performance?

Future research could investigate other confounding variables. For purposes of this study, the researcher focused on population group and gender.

5.5   METHODOLOGY

Two hundred and five students were given the same test over a period of five days. The students were required to answer 13 test items, which encompassed 58 tasks, within a timeframe of 80 minutes. They received a printout test and were asked to create tables, queries, forms and macros in Microsoft Access 2002. The students submitted their answers online on WebCT [4] where the answers were retrieved by the lecturer. The test was administered on computer in a secure, proctored environment with no teamwork or access to learning materials. The laboratory could accommodate sixty students at a time. In order to deal with any possible computer problems that might occur, a maximum of fifty students were accommodated per session. In certain sessions fewer than fifty students attended. It took one week for all 205 students to complete the test.

The demographics of the sample of 205 students were as follows:

A total of 60.2% male students and 39.8% female students participated in the research. The population group composition was 3.90% Coloured students, 10.7% White students and 85.4% Black students.

5.6   STATISTICS

For purposes of processing the data gathered, the software package SPSS was used [5]. The processing of the data includes statistics such as mean, standard deviation, and percentage. Complementary statistical techniques used were t-tests and ANOVA.

The two-sample t-test was used in order to determine whether the means of two samples differed significantly. In this study the two-sample t-test was used in an attempt to determine whether there was a significant difference between the means of males versus females in order to determine whether or not a null hypothesis could be rejected [7].

The one-way ANOVA test was used to test for significant differences between the means of population groups. This is similar to the t-test, except that more than two groups can be compared [7]. Three different population groups participated in the research, with 22 participants from the White population group, 175 participants from the Black (referred to in this study as "African") population group, and 8 participants from the Coloured population group.

In order to determine whether a hypothesis should be accepted or rejected, the probability of the statement is evaluated. If the probability is less than 0.05, then one can be sure of a conclusion. This is referred to as the level of significance (α). Another level of significance that can be used is $\alpha = 0.01$ [7]. The significance level was set at 5% ($\alpha = 0.05$).

5.6.1. One-way ANOVA test

5.6.1.1.　　　Hypotheses based on population group

5.6.1.1.1.　　Influence of population group on difficulty level

The difficulty levels of the test items (Appendix A:A1) given to students were marked as L1 (easy), L2 (medium) and L3 (difficult). In the ANOVA tests conducted, three population groups were identified, namely White, African and Coloured. The following hypotheses were formulated to determine the correlation between the difficulty level of a test item and the population group:

H_0 :　　There is no difference between the means of the three population groups in terms of the difficulty level of test items.

H_1 :　　There is a difference between the means of the three population groups in terms of the difficulty level of test items.
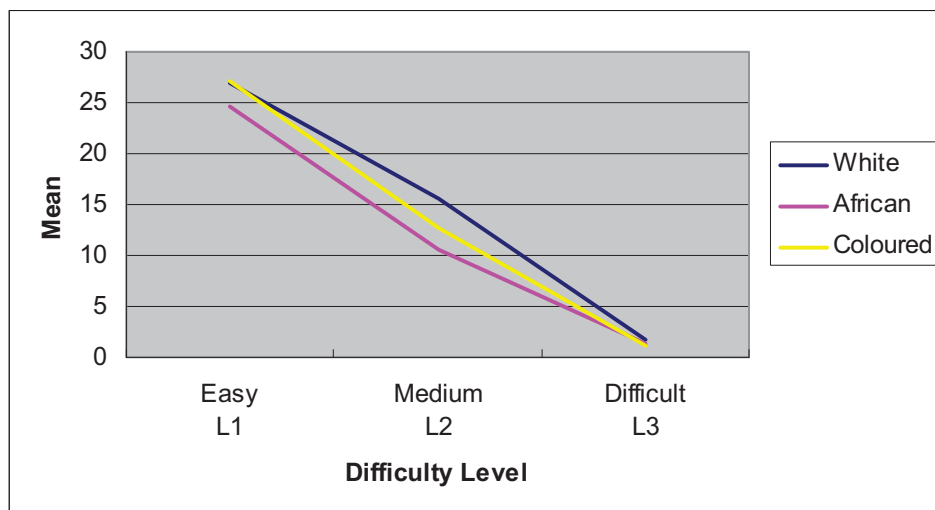
Figure 5.1: Difficulty level versus population group (mean)

The test results (Figure 5.1) for test items on difficulty level L1 (easy) that were answered correctly yielded a p-value of 0.007.　Test items answered correctly on

level L2 (medium) yielded a p-value of 0.001, while test items on level L3 (difficult) yielded a p-value of 0.254. With the significance level set at $\alpha = 0.05$, the p-value was <0.05 for test items answered correctly on difficulty levels L1 (easy) and L2 (medium). The results showed a significant difference between the three population groups with regard to test items labelled L1 (easy) and L2 (medium). H_0 is therefore rejected for the L1 and L2 test items. There is a possibility that the allocation of difficulty levels to test items on L1 (easy) and L2 (medium) levels was less accurate, while on L3 (difficult) level it was more accurate.

The hypothesis for test items answered correctly on difficulty level L3 (difficult) is accepted, as the p-value was >0.05. There were four test items marked with a difficulty level of L3 (difficult). These test items were multi-level-3 and multi-level-4 test items based on queries and forms. The students were required to create a switchboard that could open certain objects. If the switchboard could open these objects, the student received full marks, but if the switchboard could not open the specified objects, no marks were awarded. All three of the population groups performed poorly in the multi-level-4 test items marked as 'difficult'. The poor performance in this task can be ascribed to the difficulty level of the test items, as well as the fact that these items were L4 multi-level test items. Another reason could be the way in which these four items were assessed, since no marks were awarded for a partially correct task.

Table 5.1: Mean and standard deviation of difficulty level versus population group

Difficulty Level	White			African			Coloured		
	1	2	3	1	2	3	1	2	3
Mean	26.91	15.64	1.82	24.70	10.62	1.27	27.12	12.75	1.25
SD	2.47	5.56	1.74	3.75	5.96	1.42	2.36	6.52	1.28

1 = Easy, 2 = Medium, 3 = Difficult

N = 205

It was noted that the standard deviation (SD) between population groups was higher for test items at difficulty level L2 (medium) than at any other level (Table 5.1). Fifty-three percent of test items were on difficulty level L1 (easy), of which 80.78% were answered correctly. Forty percent of the test items were on difficulty level L2 (medium), of which 48.88% were answered correctly. Seven percent of the test items were on difficulty level L3 (difficult), of which 33.4% were answered correctly (Figure 5.2). It is uncertain why there was a greater flattening in the number of test items answered correctly on difficulty level L2 (medium) compared to the other two levels. One possible reason could be that some of the test items were incorrectly marked as being on difficulty level L2 (medium) while they were actually supposed to have been marked as being on difficulty level L3 (difficult).

Figure 5.2: Difficulty level versus percentage answered correctly

5.6.1.1.2.   Influence of population group on Bloom's taxonomy levels

Students received test items on Bloom's synthesis, application and analysis levels. In the statistical ANOVA, the results of the three population groups were compared against test items on Bloom's taxonomy levels (Figure 5.3).   As mentioned earlier, only test items on Bloom's application and synthesis levels were statistically analysed. The following hypotheses were formulated in order to determine the correlation between Bloom's taxonomy levels and the population group:

```
ERROR: undefined
OFFENDING COMMAND: CEBBQP+TimesNewRomanPSMT

STACK:

[/.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef
 /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef
 /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef
 /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef
 /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef
 /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef
 /zero /one /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /colon /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /
.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef /.notdef ]
/CEBBQP+TimesNewRomanPSMT*1
```