TABLE OF CONTENTS

CHAPTER 1 Introduction and background 1

1.1 Introduction 1

1.2 Background and rationale 3

1.2.1 General motivation for this study 3

1.2.2 Specific motivations 4

1.3 Problem statement and purpose of the study 6

1.4 Research questions and objectives 7

1.5 Benefits of the study 9

1.5.1 Potential contribution 9

1.5.2 Beneficiaries of the study 9

1.6 Literature study outline 10

1.7 Scope of the study 11

1.7.1 Domain and context of the study 11

1.7.2 Delimiters and limitations 11

1.7.3 Assumptions 12

1.8 Research design and methodology 13

1.8.1 Overall research approach 13

1.8.2 Literature study 15

1.8.3 Study 1: Questionnaire 1 - Identification of extent and nature of usage of e-assessment tools in South Africa 15

1.8.4 Study 2: Identification of Criteria 18

1.8.5 Study 3: Questionnaire 2 – MCQs and HOTS 18

1.8.6 Study 4: Theoretical artefact 19

1.8.7 Study 5: Electronic framework (e-SEAT) evaluation 21

1.8.8 Study 6: e-SEAT validation and application 21

1.8.9 Summary of research methods and techniques 21

1.8.10 Ethical issues 23

1.9 Structure of the thesis 24

1.10 Chapter summary 25

CHAPTER 2 Literature study: electronic assessment (e-assessment) 26

2.1 Definition of assessment 26
2.2 Types of assessment 27
2.2.1 Formative versus summative 28
2.2.2 Formal versus informal 30
2.2.3 Final versus continuous 31
2.2.4 Process versus product 31
2.2.5 Convergent versus divergent 32
2.3 Purpose of assessment 33
2.4 Measures of assessment 34
2.5 Electronic assessment (e-assessment) 34
2.5.1 Definition of e-learning 35
2.5.2 Definition of e-assessment 36
2.5.3 Features of e-assessment tools 38
2.5.4 Components in e-assessment tools 40
2.5.5 Common e-assessment tools 41
2.5.6 e-Assessment procedures 43
2.5.7 Benefits of e-assessment 45
2.5.8 Disadvantages of e-assessment 54
2.5.9 Constraints associated with e-assessment 57
2.5.10 Overcoming problems/drawbacks associated with e-assessment implementation – solutions to the constraints 59
2.6 Chapter conclusion 63
CHAPTER 3 Literature study: multiple choice questions (MCQs) in e-assessment 64
3.1 Multiple choice questions (MCQs) 64
3.1.1 Benefits of MCQs 66
3.1.2 Drawbacks associated with MCQs 67
3.1.3 MCQs for higher order thinking skills (HOTS) 69
3.1.4 Types of questions supported by e-assessment tools 70
3.1.5 Different formats of MCQs supported by e-assessment tools 71
3.2 Initial criteria for evaluating e-assessment tools/systems 87
3.2.1 Category 1: Technical criteria 88

3.2.2 Category 2: Question management criteria ... 88
3.2.3 Category 3: Test management criteria ... 90
3.2.4 Category 4: Implementation criteria ... 91
3.2.5 Category 5: Interface criteria ... 91
3.2.6 Evolution of SEAT ... 92
3.3 Chapter conclusion ... 98
CHAPTER 4 Research design and methodology ... 99
4.1 Research foundations of this study ... 100
4.1.1 Philosophical worldview ... 101
4.1.2 Worldview of present research ... 105
4.1.3 Selected strategies of enquiry ... 106
4.1.4 Strategies adopted in present study ... 111
4.1.5 Research methods ... 113
4.2 Research design ... 114
4.2.1 Action research ... 116
4.2.2 Design-based research ... 122
4.2.3 Research design selected: action research ... 126
4.3 Research methodology ... 129
4.3.1 Data collection ... 129
4.3.2 Data analysis ... 130
4.4 Frameworks ... 131
4.5 Implementation of action research in this study ... 133
4.5.1 Study 1 – Extent and nature of usage ... 136
4.5.2 Study 2 – Actual e-assessment usage ... 137
4.5.3 Study 3 – MCQ adoption patterns and HOTS ... 138
4.5.4 Study 4 – SEAT Framework ... 139
4.5.5 Study 5 – Evaluation of e-SEAT Framework ... 141
4.5.6 Study 6 – Application and validation of e-SEAT Framework ... 142
4.6 Validity, reliability and triangulation ... 143
4.6.1 Validity ... 143
4.6.2 Reliability ... 145

4.6.3 Triangulation ... 146
4.7 Chapter conclusion ... 147
CHAPTER 5 Data presentation and analysis of Phase 1 Studies ... 148
5.1 Study 1 ... 149
5.1.1 Introduction to Study 1 ... 150
5.1.2 Background to Study 1 ... 151
5.1.3 Participation: by faculty and department ... 152
5.1.4 Participation: by university ... 153
5.1.5 Quantitative findings of Study 1: users of e-assessment ... 154
5.1.6 Qualitative findings of Study 1: users of e-assessment ... 158
5.1.7 Qualitative findings of Study 1: non-users of e-assessment software ... 165
5.1.8 Conclusion and summary of Study 1 ... 169
5.2 Study 2 ... 171
5.2.1 Introduction to Study 2 ... 171
5.2.2 Interview participants ... 172
5.2.3 Quantitative findings of Study 2 ... 173
5.2.4 Qualitative findings of Study 2 ... 182
5.2.5 Conclusion and summary of Study 2 ... 190
5.3 Study 3 ... 191
5.3.1 Introduction to Study 3 ... 191
5.3.2 Questionnaire participants ... 193
5.3.3 Findings of Study 3 ... 198
5.3.4 Conclusion and summary of Study 3 ... 203
5.4 Framework evolution ... 204
5.5 Chapter conclusion ... 207
CHAPTER 6 Data presentation and analysis of Phase 2 Studies ... 209
6.1 Study 4 ... 212
6.1.1 Study 4a – Pilot Study ... 213
6.1.2 Study 4b – Evaluation Study ... 221
6.1.3 Study 4c – Proof of Concept Study ... 239
6.1.4 Study 4d – Application Study ... 244

6.1.5 Conclusion of Study 4 247
6.2 Study 5: e-SEAT Framework evaluation........ 248
6.2.1 Discussion of findings 250
6.2.2 Conclusion of Study 5 255
6.3 Study 6: e-SEAT Framework application and validation........ 255
6.3.1 Discussion of findings 257
6.3.2 Categories and criteria in e-SEAT Final Framework 261
6.3.3 Conclusion of Study 6 275
6.4 Screen prints of e-SEAT........ 276
6.5 Chapter conclusion 284
CHAPTER 7 Conclusion and Recommendations 286
7.1 Introduction........ 286
7.2 Research questions revisited........ 288
7.3 Practical and theoretical contributions of this study 299
7.4 Mixed methods research and action research 302
7.4.1 Mixed methods research........ 302
7.4.2 Action research........ 303
7.5 Validity, reliability and triangulation 307
7.5.1 Validity........ 307
7.5.2 Reliability 310
7.5.3 Triangulation........ 311
7.6 Limitations of the research........ 312
7.7 Recommendations and future research........ 313
7.7.1 Recommendations........ 313
7.7.2 Future research 313
7.8 Chapter summary and conclusion 314
REFERENCES........ 316
APPENDICES........ 331
APPENDIX A: UNISA ethical clearance documentation 331
A1: 2009 Ethical clearance for Study 1 331
A2: 2011 Addendum for Studies 2 and 3........ 331

A3: 2013 Addendum for Studies 4 to 6 ... 331
APPENDIX B: UKZN ethical clearance for Study 1 ... 331
APPENDIX C: Questionnaire 1 for Study 1 ... 331
APPENDIX D: Documents for Interviews in Study 2 ... 331
APPENDIX D1: Personal interviews schedule for Study 2 ... 331
APPENDIX D2: Informed consent for personal interviews in Study 2 (anonymized example) ... 331
APPENDIX D3: Telephonic interviews schedule for Study 2 ... 331
APPENDIX E: Questionnaire 2 for Study 3 ... 331
APPENDIX F: Documents for Study 4a - Pilot Study ... 331
APPENDIX F1: Instrument for Study 4a - Pilot Study (MS Word Version) ... 331
APPENDIX F2: Instrument for Study 4a - Pilot Study (Survey Monkey version) ... 331
APPENDIX F3: Pilot Study questionnaire ... 331
APPENDIX G: Instrument for Study 4b - Evaluation Study ... 331
APPENDIX H: Documents for Study 4c - Proof of Concept Study ... 331
APPENDIX H1: Instrument for Study 4c - Proof of Concept Study ... 331
APPENDIX H2: Proof of Concept Study questionnaire ... 331
APPENDIX I: Instrument for Study 4d - Application Study ... 331
APPENDIX J: Documents for Studies 5 and 6: SEAT and e-SEAT ... 331
APPENDIX J1: e-SEAT instrument ... 331
APPENDIX J2: SEAT questionnaire ... 331
APPENDIX J3: e-SEAT questionnaire ... 331
APPENDIX K: e-SEAT instruction file ... 332
APPENDIX L: Publications from this research ... 332
APPENDIX L1: SACLA Conference Paper ... 332
APPENDIX L2: Progressio Journal Article ... 332

LIST OF FIGURES

Figure 1.1: The data collection process ... 14
Figure 1.2: The dataset composition ... 17
Figure 1.3: Outline of study ... 22
Figure 2.1: The cycle of developing, conducting and reporting tests ... 43
Figure 2.2: e-Assessment life-cycle (Adapted from Cook & Jenkins (2010: 14)) ... 62
Figure 3.1: Multiple Choice (MCQ): single response example ... 72
Figure 3.2: Multiple Choice (MCQ): multiple response example ... 73
Figure 3.3: Extended Matching Items (EMI) example ... 75
Figure 3.4: Selection/Drop Down Lists example ... 76
Figure 3.5: True/False (T/F) example ... 77
Figure 3.6: True/False (T/F) example with the correct solution and its explanation ... 78
Figure 3.7: Fill-in-the-Blank/Completion example ... 78
Figure 3.8: Hotspot question example ... 79
Figure 3.9: Matching List example (1st year End User Computing module) ... 80
Figure 3.10: Drag-and-drop question example ... 81
Figure 3.11: Diagram question example ... 82
Figure 3.12: Simulation question example ... 83
Figure 3.13: Ranking question example ... 84
Figure 3.14: Reordering question example ... 85
Figure 3.15: Short answer question example ... 86
Figure 3.16: Evolution of the SEAT Framework ... 92
Figure 4.1: Creswell's framework for design: The interconnection of worldview, strategies of enquiry and research methods ... 100
Figure 4.2: Types of qualitative research (Merriam, 2009: 38) ... 108
Figure 4.3: Types of quantitative research ... 109
Figure 4.4: A metaphor for research design (Adapted from Mouton 2008:56) ... 115
Figure 4.5: Action research model ... 119
Figure 4.6: The cyclic process of action research ... 120
Figure 4.7: Design-based research model ... 124
Figure 4.8: Action research applied in this study ... 134
Figure 5.1: Phase 1 of the action research series ... 148
Figure 5.2: Distribution by institution, faculty and school/department ... 153
Figure 5.3: Tools currently being used ... 155
Figure 5.4: Analysis of usage of e-assessment by school/department classification ... 156
Figure 5.5: Summative vs formative usage ... 157
Figure 5.6: Reasons for non-usage ... 166
Figure 5.7: Institution participation ... 172
Figure 5.8: Number of years e-assessment adopted ... 175

Figure 5.9: Support provided for e-assessment 177
Figure 5.10: Adoption of e-assessment by type 178
Figure 5.11: Number of e-assessments per module........ 181
Figure 5.12: Computing participants per Institution 195
Figure 5.13: Non-Computing participants per Institution 195
Figure 5.14: Faculty participation 196
Figure 5.15: Usefulness of question types for HOTS – Computing users 200
Figure 5.16: Usefulness of question types for HOTS – Non-Computing users 201
Figure 6.1: Phase 2 of this research........ 210
Figure 6.2: Evolution of SEAT to e-SEAT 211
Figure 6.3: Extract from the invitation to participate in the Pilot Study 214
Figure 6.4: Sample of a screen for the Interface Design Criteria in Survey Monkey illustrating how participants had to complete the SEAT Evaluation Framework for the Evaluation Study 222
Figure 6.5: Sample of the Question Editing Criteria screen in Survey Monkey illustrating how participants had to complete the SEAT Evaluation Framework for the Evaluation Study 235
Figure 6.6: Reponses on interaction with e-SEAT: e-SEAT evaluation study........ 250
Figure 6.7: Reponses on interaction with e-SEAT: e-SEAT validation study........ 257
Figure 6.8: e-SEAT welcome screen........ 276
Figure 6.9: e-SEAT registration screen........ 277
Figure 6.10: e-SEAT post registration screen........ 278
Figure 6.11: e-SEAT Question Editing Criteria screen........ 279
Figure 6.12: e-SEAT Question Editing Criteria screen continued … 280
Figure 6.13: e-SEAT Question Editing Criteria screen continued … 281
Figure 6.14: e-SEAT report screen 282
Figure 7.1: Relationships between studies conducted and research questions posed........ 298
Figure 7.2: Components of the e-SEAT Framework........ 301

LIST OF TABLES

Table 1.1: Research questions and chapters in which they are answered 8
Table 1.2: Outline of chapters in thesis 24
Table 3.1: Framework version 1a: e-assessment criteria derived from the literature 93
Table 4.1: Creswell's philosophical worldviews (synthesised by the researcher using Creswell, 2009:6-11) 101
Table 4.2: Creswell's research methods (Creswell, 2009:15) 111
Table 4.3: Mouton's research design and methodology (Based on Mouton, 2008:56-57) 116
Table 4.4: Summary of features of design-based research model 125
Table 4.5: Summary of Study 1 as outlined in Figure 4.8 136
Table 4.6: Summary of Study 2 as outlined in Figure 4.8 137
Table 4.7: Summary of Study 3 as outlined in Figure 4.8 138
Table 4.8: Summary of Study 4 as outlined in Figure 4.8 139
Table 4.9: Summary of Study 5 as outlined in Figure 4.8 141
Table 4.10: Summary of Study 6 as outlined in Figure 4.8 142
Table 4.11: Validity in quantitative and qualitative research (Cohen et al, 2011: 182) 144
Table 5.1: Summary of Study 1 as outlined in Figure 5.1 150
Table 5.2: Distribution of participants by institution, faculty and school/department 152
Table 5.3: Tools currently adopted 154
Table 5.4: Number of years for which e-assessment had been adopted 155
Table 5.5: Question types adopted by users 156
Table 5.6: Summative vs formative usage 157
Table 5.7: Levels at which e-assessment is being used 158
Table 5.8: e-Assessment vs traditional methods of assessment 159
Table 5.9: Benefits of e-assessment outlined by adopters 160
Table 5.10: Barriers to e-assessment adoption 161
Table 5.11: Web-based e-assessment advantages 163
Table 5.12: Ease of use and accessibility 164
Table 5.13: Reasons for non-usage 166
Table 5.14: Web-based e-assessment advantages 167
Table 5.15: Non-web-based e-assessment advantages 168
Table 5.16: Summary of Study 2 as outlined in Figure 5.1 171
Table 5.17: Tool embedded in the Learning Management System (LMS) 174
Table 5.18: Tools adopted 175
Table 5.19: Change in tool adoption 176
Table 5.20: Policies on the adoption of e-assessment 176
Table 5.21: Contribution of e-assessment to final mark 177
Table 5.22: Mark release 179
Table 5.23: Levels where e-assessment is adopted 180

Table 5.24: Number of e-assessments per module 181
Table 5.25: Formal training on e-assessment 182
Table 5.26: Administration of e-assessment 182
Table 5.27: Benefits of e-assessment 183
Table 5.28: Disadvantages of e-assessment 187
Table 5.29: Summary of Study 3 as outlined in Figure 5.1 193
Table 5.30: Participant distribution by institution 194
Table 5.31: Participant distribution by school/department 197
Table 5.32: Participant distribution by position held 197
Table 5.33: Average adoption of question types – comparison of groups 198
Table 5.34: Suitability of questions per level of study 202
Table 5.35: Framework version 1b - e-assessment criteria identified in the empirical studies 205
Table 6.1: Summary of Study 4a as outlined in Figure 6.1 213
Table 6.2: Question themes for pilot participants 216
Table 6.3: Changes to Pilot Framework based on pilot participants' reviews, Study 4a 219
Table 6.4: Summary of Study 4b as outlined in Figure 6.1 221
Table 6.5: Statistical Analysis of the Interface Design Criteria 224
Table 6.6: Statistical Analysis of the Question Editing Criteria 225
Table 6.7: Statistical Analysis of the Assessment Strategy Criteria 226
Table 6.8: Statistical Analysis of Test and Response Analysis Criteria 227
Table 6.9: Statistical Analysis of Test Bank Design Criteria 228
Table 6.10: Statistical Analysis of Security Criteria 229
Table 6.11: Statistical Analysis of Compatibility Criteria 230
Table 6.12: Statistical Analysis of Ease of Use Criteria 231
Table 6.13: Statistical Analysis of Technical Support Criteria 232
Table 6.14: Statistical Analysis of Question Types Criteria 233
Table 6.15: Statistical Analysis of Test and Response Analysis Criteria 236
Table 6.16: Statistical Analysis of Ease of Use Criteria 236
Table 6.17: Changes made to SEAT Framework after the main evaluation in Study 4b 238
Table 6.18: Summary of Study 4c as outlined in Figure 6.1 240
Table 6.19: Recategorisation of criteria in the SEAT Framework 241
Table 6.20: Initial rating scale used in the SEAT Framework (Pilot, Evaluation and Proof of Concept Frameworks) 241
Table 6.21: Adapted rating for the SEAT Framework after Study 4c 241
Table 6.22: Changes made to SEAT Proof of Concept Framework after Study 4c 242
Table 6.23: Summary of Study 4d as outlined in Figure 6.1 244
Table 6.24: Summary of Study 5 as outlined in Figure 6.1 248
Table 6.25: Aspects discussed in Study 5 250
Table 6.26: Summary of changes made to e-SEAT based on the results of Study 5 254
Table 6.27: Summary of Study 6 as outlined in Figure 6.1 256

Table 6.28: Aspects discussed in Study 6 ... 257
Table 6.29: Summary of changes made to e-SEAT Validation Framework based on the results of Study 6 ... 260
Table 6.30: Question Editing category in e-SEAT ... 262
Table 6.31: Question Editing category in e-SEAT continued ... 263
Table 6.32: Assessment Strategy category in e-SEAT ... 264
Table 6.33: Test and Response category in e-SEAT ... 265
Table 6.34: Test and Response category in e-SEAT continued … ... 266
Table 6.35: Test Bank category in e-SEAT ... 267
Table 6.36: Question Types category in e-SEAT ... 268
Table 6.37: Interface Design category in e-SEAT ... 269
Table 6.38: Security category in e-SEAT ... 270
Table 6.39: Compatibility category in e-SEAT ... 271
Table 6.40: Ease of Use category in e-SEAT ... 272
Table 6.41: Robustness category in e-SEAT ... 273
Table 6.42: Technical Support category in e-SEAT ... 274
Table 7.1: Research questions and chapters in which they are answered ... 288
Table 7.2 Overview of e-SEAT Framework criteria evolution ... 306
Table 7.3: Validity in this research ... 307
Table 7.4: Reliability in this research ... 310
Table 7.5: Triangulation in this research ... 311

CHAPTER 1 Introduction and background

This chapter commences with an introduction to the research, in Section 1.1. Section 1.2 focuses on the background to the study, including a general motivation for the research (Section 1.2.1) and the specific motivation (Section 1.2.2). The problem statement and research questions are presented in Sections 1.3 and 1.4 respectively. The benefits of the study (Section 1.5) are discussed under two subsections – namely the potential contribution of this study (Section 1.5.1) and the beneficiaries of the study (Section 1.5.2). A brief outline of the literature study is presented in Section 1.6. The scope of the study is discussed in Section 1.7 with Section 1.7.1 focusing on the domain and context, Section 1.7.2 presenting the delimiters and limitations, and Section 1.7.3 discussing the assumptions. Section 1.8 is a comprehensive discussion on the research design and methodology, discussing the approaches used in each of the substudies of this research. Section 1.9 outlines the structure of the thesis and is followed in Section 1.10 by a summary of the chapter.

1.1 Introduction

The time-independent, space-independent and location-independent nature of Information and Communication Technology (ICT) can provide 'decongestion of overcrowded education facilities, support for students and educators, and a valuable opportunity for specific groups of students if the learning material is accessible to them' (Ardito, Costabile, de Marsico, Lanzilotti, Levialdi, Roselli & Rossano, 2006: -12). Cox (2013) suggests that online learning enables students to study anywhere and at any time. This applies not only to direct instruction and learning, but is also relevant to assessment, in the form of tests and examinations. e-Assessment is a domain of e-learning where Information and Communication Technology (ICT) is used to administer assessments and record students' responses (Dube and Ma, 2011). Research and development have led to the development of various e-assessment tools and systems that can create, deliver, mark, analyse and provide customised online and paper-based reporting services for both summative and formative assessments (Harrington & Reasons, 2005).

The purpose of this research is to develop, validate, apply and refine a framework for the evaluation of e-assessment tools being used, or under consideration for adoption, at higher-education institutions in South Africa. The particular environment of this study is schools and departments in Computing-related disciplines, namely: Computer Science (CS), Information Systems (IS) (also termed Informatics), and Information Technology (IT).

The primary aim of this study is to undertake iterative development and research on such a framework, comprising several categories of criteria that can be used by educators to assist them in the selection of an e-assessment tool (interchangeably termed an e-assessment system), to facilitate the adoption of electronic assessment at their institutions. This framework can also be used to evaluate e-assessment systems and tools already in use. Furthermore, the criteria included in the framework also serve as design guidelines for designers creating new systems.

The secondary aim of this study is to understand the current extent and nature of use of e-assessment tools, as well as the satisfaction afforded to the users, namely the academics. The study is centred on local usage in South Africa and, to a small degree, considers international usage. Although this was a secondary aim, this work on the nature and extent of use was conducted at an early stage, since its findings were used in the development of the framework.

The study uses:

- literature, based mainly on international experience,
- data, regarding local experiences with e-assessment in South Africa, and
- quantitative and qualitative studies, to
    - gather information on the current extent and nature of usage of e-assessment tools, as well as levels of satisfaction with such tools,
    - identify the types of questions commonly adopted in e-assessment in South Africa,
    - understand the role of multiple choice questions (MCQs) in testing higher order thinking skills (HOTS),
    - develop a framework to
        - assist users in the selection and acquisition of e-assessment tools,
        - provide design guidelines for developers, and
    - evaluate, refine and apply the evaluation framework developed through a series of action research studies.

In Phase 1 of this study, the local adoption of e-assessment systems and the types of questions these systems supported were investigated in a series of studies. Upon completion of Phase 1, the evaluation framework named SEAT (Selecting and Evaluating an e-Assessment Tool), was developed in Phase 2 in an iterative manner, which involved creating, evaluating and refining the framework. Thereafter, SEAT was converted to an electronic evaluation framework, named e-SEAT (electronically Selecting and Evaluating an e-Assessment Tool) through a similar series of development, evaluation and refinement. SEAT was developed both from literature studies and from the findings of user-based surveys, via questionnaires and interviews. The e-SEAT Framework was iteratively validated during its evolution, by further questionnaire and interview research, the findings of which were used to refine it. Finally it was applied to evaluate various systems used for assessment at tertiary institutions in South Africa.

Disclaimer: This research is situated in the subdiscipline of multiple choice question (MCQ)-related assessment. The author acknowledges the subdiscipline of automated assessment whereby essay-style questions are judged using sophisticated techniques of pattern matching, natural language processing, and artificial intelligence. However, these forms of assessment are outside the scope of the present study.

1.2 Background and rationale

1.2.1 General motivation for this study

The present research was initiated by:

- the researcher’s personal interest in the area of e-assessment,
- the researcher’s motivation for the aims to be achieved, and
- suggestions in the literature that this is a domain where more research is required (Christakoudis, Androulakis & Zagouras, 2011; Deutsch, Herrmann, Frese & Sandholzer, 2012; Pretorius, Mostert & de Bruyn, 2007; Valenti, Cucchiarelli & Panti, 2002; Yonker, 2011).

The aspects briefly outlined in this subsection are elaborated in Section 1.2.2.

1.2.2 Specific motivations

As first-level Information Systems co-ordinator for seven years at the University of KwaZulu-Natal (UKZN), the researcher faced a major challenge regarding the selection and implementation of an e-assessment tool for student assessment. Due to the large number of students, approximately 1600 students, distributed across two campuses, the method of assessment adopted in the School of Information Systems and Technology (IS&T), for its entry-level students, was primarily in the family of MCQs. Since 2003, the researcher has experimented with various e-assessment tools for judgement of MCQs, including: SAM (Cengage Course Technology samcentral.course.com), ExamView Test Generator (Pearson Assessments http://www.formative-assessments.com/formative/examview/index.htm), Hot Potatoes (Half-Baked Software Inc. hotpot.uvic.ca), EzTest Online (McGraw Hill www.eztestonline.com) and CourseCompass (Pearson Assessments www.coursecompass.com). Various problems were encountered during the testing of these e-assessment software tools, two of the major ones being:

- tool interfaces were not easy for students to understand, and
- the administration associated with implementing the tools was laborious.

Similar disabling factors are encountered at other South African tertiary institutions (Brink & Lautenbach, 2012; Researcher Interviews, 2009 to 2012; Singh & de Villiers, 2010). Respondents indicated satisfaction with e-assessment, particularly when conducting assessments with larger classes, by using multiple choice questions that can be assessed automatically. They encountered problems, however, with access to e-assessment tools, which was a barrier to their long-term implementation. Furthermore, financial constraints, lack of infrastructure and poor commitment by senior academics, made the use of e-assessment complex for interested academics. Due to pressurised schedules, heavy tuition loads and research commitments, academics lack the time to initiate new ventures or to investigate the new range of possible technologies available to them. Systems that provide support in investigating and adopting new approaches to assessment would save time in the longer term. Thus the development of an evaluation framework that would assist academics in the decision of which e-assessment tool to adopt, is essential to their making more informed decisions.

In the literature, mention is made of the need for increased research and further studies on the use of e-assessment and electronic testing systems (Christakoudis, Androulakis & Zagouras, 2011; Derczeni & Rogozea, 2011; Deutsch, Herrmann, Frese & Sandholzer, 2012; Gilles, Detroz & Blais, 2012; Pretorius, Mostert & de Bruyn, 2007; Valenti, Cucchiarelli & Panti, 2002). Derczeni and Rogozea (2011) indicate that e-assessment is being adopted more frequently in the evaluation of students' knowledge. However, although there is increased interest in, and adoption of, e-assessments in higher education, for successful implementation of e-assessment, students' attitudes and reservations should be researched (Deutsch, Herrmann, Frese & Sandholzer, 2012). Laumer and von Stetten (2009) believe that the benefits of e-assessment are two-fold – that is, they can provide support for existing educational goals, as well as assist in assessing students' problem-solving and processing skills. Gilles, Detroz and Blais (2012) state that the effectiveness and efficiency of e-assessment should be investigated to determine if there are any positive links between teaching, the quality of learning and the modes of learning assessment favoured. A concern raised by Christakoudis, Androulakis and Zagouras (2011) is that, despite e-assessment systems being widely adopted, they often do not cater well for the repeated use of the questions stored in their question banks, although many of the systems support automatic question randomisation.

Following the failure to successfully implement an e-assessment tool in the School of IS&T at UKZN, yet taking cognisance of recent advances in educational technologies, the researcher set out to conduct this research and generate an evaluation framework to support educators in the evaluation and selection of appropriate e-assessment tools for adoption. As a first step, an investigation was conducted to assess the current extent and nature of usage of e-assessment tools within Information Systems (IS), Information Technology (IT) and Computer Science (CS) academic departments and schools at South African tertiary institutions, so as to determine the levels of satisfaction or dissatisfaction experienced by academics in acquiring and using these technologies. The findings of this initial localised identification-of-use study, as described in Study 1 (Section 5.1), set the context for the rest of the research.

1.3 Problem statement and purpose of the study

The problem that underlies this study is the lack of a comprehensive and readily-available means of evaluating e-assessment tools that administer MCQs. There is a need for an evaluation framework that would support academics in evaluating MCQ tools being considered for adoption or MCQ systems already in use in their institution.

Despite widespread and increasing international usage of e-assessment tools (Buchan & Swann, 2007; Feng, Heffernan & Koedinger, 2009; Hasibuan & Santoso, 2005; Hodson, Saunders & Stubbs, 2002; Honarmand, 2009; Kadhi, 2004; Khedo, 2005; Laborda & Royo, 2008; Messing, 2004; Moskal, Dziuban, Upchurch, Hartman & Truman, 2006; Testa 2008), South African (SA) universities are not adequately implementing this form of assessment. As such, there is scope for further use to realise the full potential and benefits of e-assessment (Boyle & Hutchinson, 2009; Brink & Lautenbach, 2012; Singh & de Villiers, 2010).

The *focus area* of this study is therefore the design of a framework to be used for the evaluation of e-assessment tools that assist to automate the assessment of questions in the MCQ family. The *application area* is the domain of e-assessment within Computing-related academic departments and schools at South African tertiary institutions.

Specifically, and in more detail, the research sets out to:

- establish the extent and nature of current usage in South Africa, as well as levels of satisfaction with such tools (Study 1 in Section 5.1 and Study 2 in Section 5.2). Study 2 also contributed to identify users' requirements in e-assessment systems,
- determine the types of e-assessment questions being adopted in South Africa, and their role in testing higher order thinking skills (HOTS) (Study 3 in Section 5.3),
- investigate requirements for e-assessment systems and associated categories of criteria for evaluation, applicable to the South African situation (Study 4 in Section 6.1). This was achieved through an iterative process of data collection which commenced with a pilot study, evolving into evaluation, application and validation studies respectively, and

- develop an evaluation framework to facilitate the evaluation, adoption, and design of e-assessment tools in South Africa (Study 5 in Section 6.2 and Study 6 in Section 6.3). This framework evolved from a manual framework named SEAT (Selecting and Evaluating an e-Assessment Tool) to an electronic framework named e-SEAT (electronically Selecting and Evaluating an e-Assessment Tool), during the iterative studies in Phase 2.

The target group of participants primarily comprises local South African academics in Computing-related departments who are either current or potential users of e-assessment tools.

1.4 Research questions and objectives

The main research question for this study is:

"How does an academic evaluate an e-assessment tool, to identify the best-fit for his/her requirements?"

The associated objective of this research is to develop a framework that facilitates the evaluation of e-assessment systems.

This section lists the research questions addressed in the study (see Table 1.1), followed by a brief outline of the purpose of the envisaged framework. The intended group of participants is also described.

Table 1.1: Research questions and chapters in which they are answered

	Research question	**Chapter(s) in which answered**
RQ 1	What is the extent and nature of use of electronic assessment in Computing-related departments in South African universities?	5 (Section 5.1 and Section 5.2)
RQ 2	What types of questions are being adopted in e-assessment systems in South Africa?	5 (Section 5.3)
RQ 3	How appropriate are these questions (identified in Research Question 2) for testing higher order thinking skills (HOTS)?	5 (Section 5.3.3)
RQ 4	What are the requirements for selecting or personally developing an electronic assessment tool? • Theory: What does the literature suggest as appropriate requirements for electronic/online testing and assessment tools? • Practice: What criteria are used in practice in South African higher institutions for the selection and use of electronic/online testing and assessment tools?	3 (Section 3.2.6; Table 3.1 Table 3.1) 5 (Section 5.4; Table 5.35 Table 5.35)
RQ 5	What categories and criteria should be incorporated in a prototype framework to evaluate electronic assessment systems?	6 (Sections 6.1.1; 6.1.2; 6.1.3)
RQ 6	How appropriate and effective is the proposed framework?	6 (Sections 6.1.4; 6.2; 6.3)

The over-arching objective of the study is therefore to garner a set of criteria for the envisaged evaluation framework. These criteria were derived both from the literature (secondary data in Section 2.2) and from research (primary data in Sections 5.2 and 5.3). As stated in Section 1.1, this framework will serve a dual purpose in that it will present categories and criteria for evaluating existing e-assessment systems, and provide design guidelines for the development of new such systems.

To this end, the research was primarily undertaken in IS, IT and CS departments, aiming to answer the research questions outlined in Table 1.1. However, certain Non-Computing academics and international academics who could make relevant contributions to the research also participated. These external participators were invited on a basis of convenience sampling and snowball sampling, when the researcher was referred to them by South African academics or through references in the literature. In processing the data, however, it will explicitly be made clear from which group the participants come (Computing academic or Non-Computing academic). The data from international participants was combined with the Non-Computing academic data set due to the small number of participants. The findings, as required, are presented separately, as well as in a consolidated form (see Figure 1.2 in Section 1.8.3).

1.5 Benefits of the study

This section presents the potential contributions associated with this study, and also outlines the beneficiaries of this study.

1.5.1 Potential contribution

As stated, the major outcome of this study is the development of an evaluation framework to facilitate the evaluation, adoption, and implementation of e-assessment in South Africa.

Further contributions of this study include:

- adding to the body of knowledge on e-assessment by providing a structured set of categories and criteria for evaluating e-assessment systems,
- building on the literature available with regard to questions of the MCQ family,
- deepening the understanding of the implementation and usage of e-assessment tools, and
- providing design guidelines to developers of e-assessment tools.

1.5.2 Beneficiaries of the study

The findings from this research will be useful to:

- academics at higher educational institutions who wish to implement e-assessment software for assessment purposes,
- designers and developers of testing applications, who will be able to incorporate the requirements and criteria identified by this research as design guidelines to facilitate the development of new products,

- students, who will benefit from the use of these new educational technologies as a replacement for, or a supplement to, traditional methods of assessment, and
- educational institutions, which will benefit from increased research productivity of their academics, due to saving time previously spent on manual assessments.

1.6 Literature study outline

The study consists of two literature reviews, with

- Chapter 2 focusing on literature on assessment in general (Sections 2.1 to 2.4), leading on to e-assessment (Sections 2.5 and 2.6), and
- Chapter 3 discussing the literature available on MCQs in e-assessment (Section 3.1) as well as detailing criteria used when selecting and adopting e-assessment tools (Section 3.2).

Chapter 2 presents definitions and concepts associated with assessment. It also outlines various types of assessment identified in the literature (Section 2.2), such as formative, summative, convergent and divergent assessments. The chapter also discusses the purpose (Section 2.3) and measures (Section 2.4) of assessment. Thereafter, the discussion moves to e-assessment (Section 2.5) – its definition; features, procedures, benefits, disadvantages and constraints.

Chapter 3 commences with the discussion of the MCQ genre (Section 3.1) specifically those types and formats adopted in e-assessment tools in Sections 3.1.4 and 3.1.5 respectively. Chapter 3 further outlines an initial range of criteria for selecting and adopting e-assessment tools as discussed in the literature. They are grouped into the following categories:

- Technical criteria in Section 3.2.1
- Question creation and management in Section 3.2.2
- Test management in Section 3.2.3
- Implementation in Section 3.2.4, and
- Interface in Section 3.2.5.

These five categories are based on those created by Pretorius, Mostert and de Bruyn (2007), and Valenti, Cucchiarelli and Panti (2002). Following further literature studies, the researcher subsequently adapted and expanded these five categories into ten categories to facilitate the creation of a comprehensive evaluation framework.

1.7 Scope of the study

1.7.1 Domain and context of the study

- This research focuses on establishing a framework to evaluate e-assessment systems under consideration for adoption in the over-arching domain of higher education in Information Systems (IS) and other Computing-related disciplines.
- Existing international literature and data from the experience of participants form the foundations of this study, thus setting the context and creating a general frame of reference.
- A variety of categories is established for structuring the framework of criteria.
- Participants are users and potential users of e-assessment tools, mainly from Computing-related departments.

1.7.2 Delimiters and limitations

- The research is aimed at supporting South African academics and is conducted mainly within South Africa, but participants also include international academics who volunteered, or who were referred to the researcher, or who were identified from the literature and requested by the researcher to participate.
- The application area is restricted to higher educational institutions.
- The investigation of multiple choice questions includes several variants of questions within the MCQ genre (for example, fill-in-the-gap, matching columns, hotspots on diagrams), as well as short-answer questions with limited pattern-matching capabilities.
- The context of the evaluation framework is restricted to the more common forms of assessment and testing identified in Study 1 (Section 5.1), the identification-of-use survey.
- As stated in the Disclaimer in Section 1.1, the use of artificial intelligence techniques for analysing textual responses is outside the scope of the present research. Similarly, the study does not investigate the use of text analysis tools, such as those implemented by sophisticated pattern matching techniques and natural language processing.

- A basic interactive electronic instrument, named e-SEAT, has been developed, on which to implement the SEAT Framework. This prototype e-SEAT has limited functionality and should not be viewed as an operational system for public use.
- This research does not investigate students' perspectives on e-assessment.
- The focus of this research is restricted to the application of e-assessment in controlled testing environments.
- The framework developed as a result of this research is an evaluation framework and not a conceptual framework (Section 4.4).
- The following terms are used interchangeably in this study:
  - 'participant' and 'respondent',
  - 'e-assessment tool' and 'e-assessment system',
  - 'SEAT instrument' and 'SEAT tool',
  - 'e-SEAT instrument' and 'e-SEAT tool',
  - 'survey' and 'questionnaire', and
  - 'academic' and 'educator'.

1.7.3 Assumptions

- A fundamental underlying assumption of the study is that the participants had an understanding of the nature and purpose of an e-assessment system.
- It is assumed that the participants had an adequate command of English, which is the most common language used in e-assessment systems.
- It is assumed that the questionnaires were completed by the intended persons and that such participants provided authentic and honest opinions.

1.8 Research design and methodology

1.8.1 Overall research approach

The research approach comprises literature studies and six empirical studies conducted over a period of four years. The underlying research design is action research which, as Elliott (1991) states is also termed participatory research. Action research, described in Section 4.2.1, involves a series of cycles which include planning, observing, reflecting, then re-planning, acting and observing. This was achieved through a series of six main studies and four substudies, which followed the iterative nature of action research through development, evaluation, application and refinement of e-SEAT. The strength of action research lies in its focus on generating solutions to a practical problem. In this case, the action research aimed at developing an evaluation framework as a solution to the practical issue that academics face in selecting an appropriate e-assessment tool for implementation.

The series of studies in this research have been approved by the Ethical Clearance Committee of the College of Science, Engineering and Technology (CSET) of the University of South Africa (UNISA) (see Appendices A1, A2 and A3) while the initial study, Study 1, was also approved by the Ethical Clearance Committee of the University of KwaZulu-Natal (UKZN) (see Appendix B). The subsequent studies carried out in this research did not require further Ethical Clearance by UKZN.

Figure 1.1 summarises the research and data collection processes of the various studies. The research techniques included literature reviews, interviews, questionnaires, and observations. As explained in Section 1.4, the participants were mainly from IS, IT and CS departments or schools.

Phase 1 comprised three studies, namely Studies 1, 2 and 3 while Studies 4, 5 and 6 made up Phase 2 of this research. The first version of the framework, termed 'artefact' in Figure 1.1, was developed between Study 3 and Study 4. The figure also shows how the evaluation criteria identified from the literature, and from data obtained in Studies 2 and 3, were combined prior to Study 4.

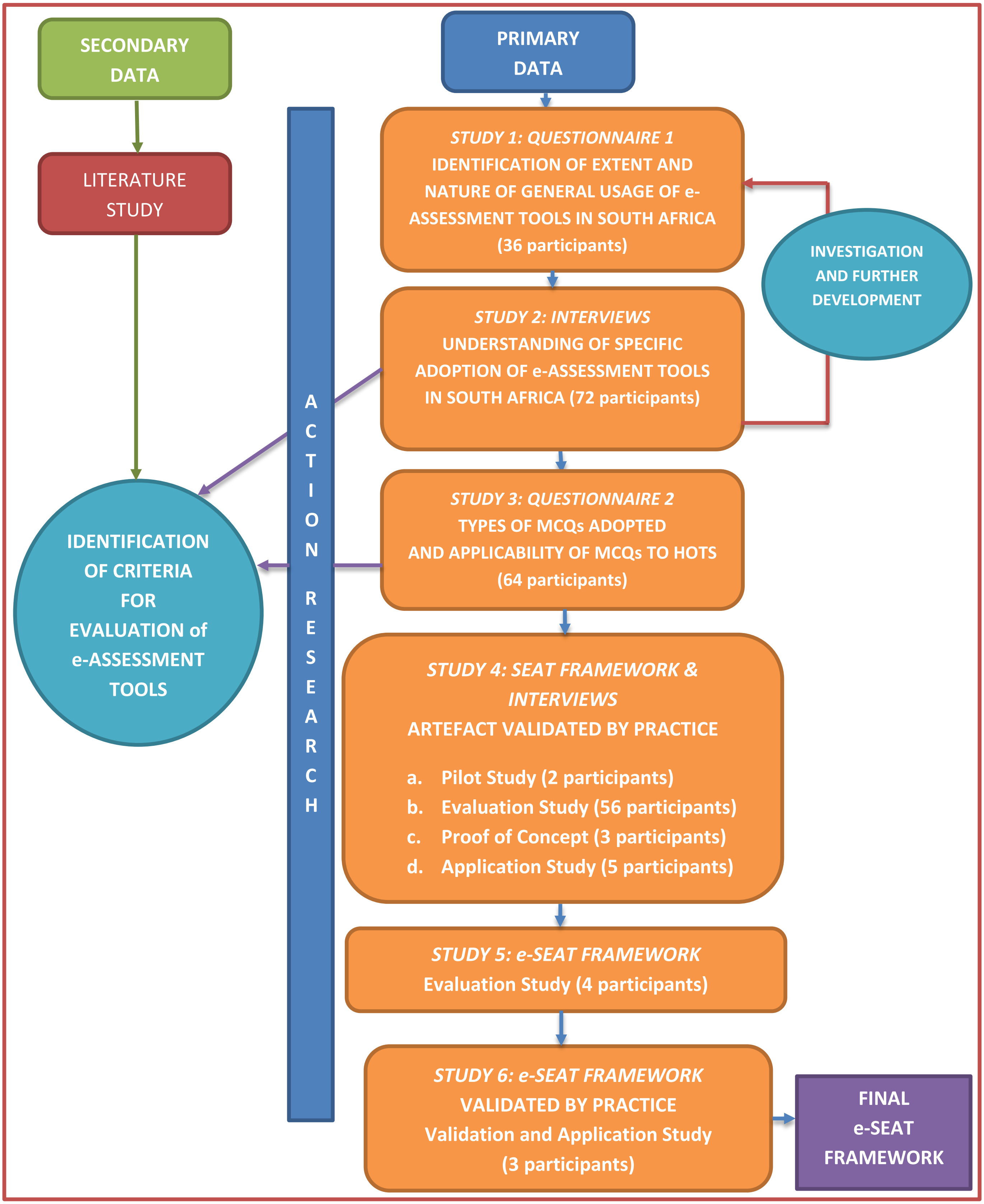


Figure 1.1: The data collection process

1.8.2 Literature study

The methodologies adopted in this research involve literature studies to obtain a conceptual background of the research area and to provide secondary data. Through a literature study, the researcher was able to 'decide upon viable research questions that have not been fully addressed' by the current literature sources (Oates, 2010: 34; Miles & Huberman, 1994). A review of the literature is essential, because in this way:

- duplication of previous studies is avoided;
- an understanding is developed of the most 'recent and authoritative theorising about the subject';
- the most widely accepted empirical findings are understood;
- the most widely accepted definitions of key concepts are ascertained (Mouton, 2008: 87).

Information for the literature studies was acquired using the following resources:

- personal keyword searches via internet-based search engines and library electronic database searches;
- identification of key references listed at the end of research articles and personally accessing these publications;
- consultations with librarians at UKZN and UNISA;
- review of existing theses in related domains.

1.8.3 Study 1: Questionnaire 1 - Identification of extent and nature of usage of e-assessment tools in South Africa

The first set of data in the action research series was obtained by identifying the 'extent and nature of use' of assessment tools in the context of South African computing education, through Questionnaire 1 (see Appendix C). In cases where established e-assessment and/or automated testing policies existed, participants were academics, appropriate members of management, or test facilitators. In cases where only ad hoc use occurred, often initiated by individual 'champions', participants were the relevant educators. All Computing-related departments in South Africa (SA) were invited to participate in the study. Gatekeeper consent was required from those departments and schools that agreed to participate. Once gatekeeper permission had been obtained, the questionnaire was mailed to the participating academic departments, where the researcher requested that it should be announced or loaded the electronic notice board or any other forum. In practice, on several occasions, initial access to a department or school occurred via an individual member of staff who was involved in e-assessment or who was aware of a colleague doing so.

The purpose of Study 1 was to 'add to what is known about the specialist subject, through a literature based survey' (Oates, 2010: 17, 23) by adding new real-world data regarding e-assessment usage in South Africa to supplement information from the literature.

In Study 1, the empirical research took the form of a user-based questionnaire (see Appendix C), where the respondents were the educators/academics who used, or who had considered using, e-assessment. In order to gain an overall perspective on the extent of usage of e-assessment, as well as on the nature of such use, the questionnaire investigated the number of years and levels at which e-assessment had been used; types of tools adopted by the participants; types of questions supported by the tools; usage of tools for summative and formative assessment; and respondents' views on the debate surrounding the use of online versus traditional assessment.

Items for the questionnaire were developed from concepts encountered in the literature. There was a predefined set of questions, with alternative sections to be completed, depending on the participants' level of usage of e-assessment. The survey results provided the researcher with data for analysis and interpretation. This data helped to identify the current situation with regard to e-assessment in CS, IS and IT academic departments at tertiary institutions in South Africa, and constituted the first set of primary data for the overall study.

Although the target group of Study 1 was academics in South African Computing schools and departments, the researcher was aware that usage, at this stage, was limited to a small minority. Where further participants became available due to referrals and personal networking, they were included in the study, but their separate affiliations were clearly indicated. In other words, Study 1 was conducted as a living study and the data was processed iteratively as more became available.

As depicted in Figure 1.2, the groups of participants consisted of:

- Academics from South African Computing-related departments and schools, including Computer Science, Information Systems, Information Technology, Information Science and e-Learning educators, as indicated in Dataset 1.
- Academics from South African Non-Computing departments and schools, as well as international participants. These are listed in Dataset 2.

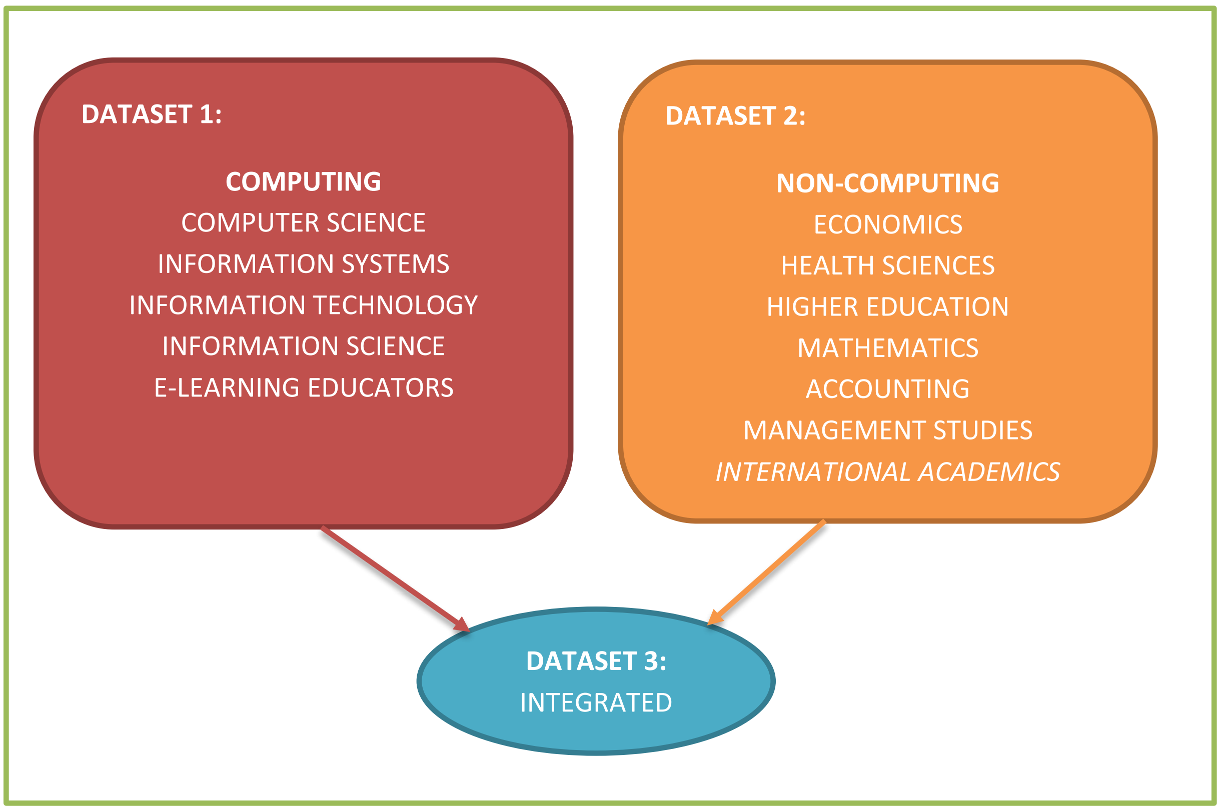


Figure 1.2: The dataset composition

Dataset 3 is thus an integrated set comprising Dataset 1 and Dataset 2. The findings of Study 1 are presented in Section 5.1.

1.8.4 Study 2: Identification of Criteria

In Study 2, in-depth personal interviews were conducted with a much larger sample of selected academics, in order to better understand usage of the various e-assessment systems currently adopted in South Africa. Eighteen of the 72 interviewees were respondents who had participated in Study 1. The remaining 54 interviewees, who had not participated in Study 1, were acquired by referrals and personal networking.

For the latter group of interviewees, that is, those who had not been involved in Study 1, the basic set of interview questions (see Appendix D1) was augmented by adding questions (see Appendix D3) that were part of the initial Study 1. The objective of the interviews in Study 2 was to extend the groundwork provided by the literature, and this interaction assisted the researcher in establishing the interviewees likes, dislikes and requirements, which, in turn, contributed to a set of criteria for the framework to be developed. The findings of Study 2 are presented in Section 5.2.

1.8.5 Study 3: Questionnaire 2 – MCQs and HOTS

Another questionnaire, Questionnaire 2 (see Appendix E), was used in Study 3. It focused on obtaining respondents' opinions on the types of multiple choice questions they prefer to, or actually, use.

This short questionnaire was used for Study 3, to ascertain information about academics' views on the different types of MCQs that can be supported in an e-assessment system. Furthermore, a section in Questionnaire 2 investigated how applicable these MCQ types are to testing higher order thinking skills (HOTS).

The data collected from Study 3 was fed into Study 4 – development of the framework – as a subcategory on the types of questions to be included in an e-assessment tool. This subcategory thus contributed to identification of criteria based on users' requirements. The findings of Study 3 are presented in Section 5.3.

Phase 1 of the action research series comprised Studies 1, 2 and 3, which contributed towards building the theoretical foundation, shown in Figure 4.8, in the chapter on research design and methodology.

1.8.6 Study 4: Theoretical artefact

Based on the foundation laid in Phase 1, Phase 2 consists of Studies 4, 5 and 6, which address the creation and refinement of the evaluation framework.

The next set of data in the action research series related to the generation of a set of criteria that could be used for the development of a framework for evaluating e-assessment systems. The reviewed literature served as the main source of concepts that supported the researcher in constructing appropriate categories and criteria for the evaluation framework.

In addition, Study 2 and Study 3 provided valuable data from practice and experience regarding the criteria viewed by participants as being important in the selection, use, and evaluation of such applications. This information was obtained from educators, managers and designers who are actual users or stakeholders of e-assessment. Study 4 aimed to develop/synthesise a comprehensive evaluation framework using both the literature and input from peers who have expertise and experience with e-assessment systems.

Studies 1, 2 and 3 were theoretical and conceptual. Study 4, by contrast, was practical, as it iteratively developed, investigated and validated the SEAT Framework generated from the findings from the literature, Study 2 and Study 3. This development process involved four substudies: Study 4a, Study 4b, Study 4c and Study 4d, which are overviewed in subsections 1.8.6.1, 1.8.6.2, 1.8.6.3 and 1.8.6.4 respectively.

1.8.6.1 Study 4a – Pilot Study

A small sample of academics, who were colleagues of the researcher, were asked to critically evaluate the initial version of the framework developed for evaluating e-assessment tools. Their comments, changes and concerns, as reported in this Pilot Study, were addressed and the researcher developed a refined instrument as a prototype framework for Study 4b.

1.8.6.2 Study 4b – Evaluation Study

Academics who had participated in Study 1 and in Study 2, as well as some others whom they recommended, were asked to test the utility of the prototype framework that emerged from Study 4a by performing a heuristic evaluation on it in an extensive Evaluation Study. As already stated, the main purpose of this stage of the process was to validate and refine the emerging categories and criteria of the framework.

1.8.6.3 Study 4c – Proof of Concept Study

The investigation done by participants in this Proof of Concept Study was similar to that of Study 4b. However, the participants were a small and select sample of leading experts in the field of e-assessment, who were invited to critically evaluate the framework to confirm its utility, as modified after the findings of Study 4b.

1.8.6.4 Study 4d – Application Study

As in Study 4c, participants in this Application Study were also an invited sample of leading experts in the field. They were selected because they are among South Africa's greatest and most experienced users of e-assessment. They were given the resulting framework, called SEAT (Selecting and Evaluating an e-Assessment Tool), as it emerged, refined, from Study 4c. They were each required to 'try it out' by applying it to an existing e-assessment system, which they had previously adopted or intended adopting in the future.

Investigation and evaluation of the e-assessment applications themselves, was a secondary contribution of this research. The names of the systems so evaluated will not be disclosed in this thesis, although the findings regarding these e-assessment tools are available for interested stakeholders.

The findings of all the substudies of Study 4 are discussed in Section 6.1.

1.8.7 Study 5: Electronic framework (e-SEAT) evaluation

Study 5 in this action research process, involved conversion of the final version of the SEAT Framework, as it emerged from Study 4, to a prototype electronic framework, named e-SEAT (electronically Selecting and Evaluating an e-Assessment Tool), that can be used by both management and academics who are keen on evaluating or adopting e-assessment tools at their institution.

In Study 5, the e-SEAT Framework was taken to a selected group of participants for an Evaluation Study. Whereas Study 4b was an evaluation of the 'manual' version of the SEAT Framework, Study 5 was an evaluation of the electronic version. Section 6.2 presents the findings of Study 5. After evaluation of this version of e-SEAT, the ultimate study, Study 6, was undertaken to finalise the prototype that emerged from the action research process.

1.8.8 Study 6: e-SEAT validation and application

Study 6 was the ultimate process in this action research series. The participants were selected from those identified as key users of e-assessment systems and were thus invited to participate in this final study in the action research process. Validation and application of the e-SEAT Framework was completed in this study, the findings of which are given in Section 6.3. This resulted in the ultimate product of this PhD Study, the Final e-SEAT Framework.

The e-SEAT Framework also serves as a set of design guidelines for designers developing online testing systems.

1.8.9 Summary of research methods and techniques

Figure 1.3 outlines and summarises the research methods of the entire research process, as described in Sections 1.8.1 to 1.8.8.

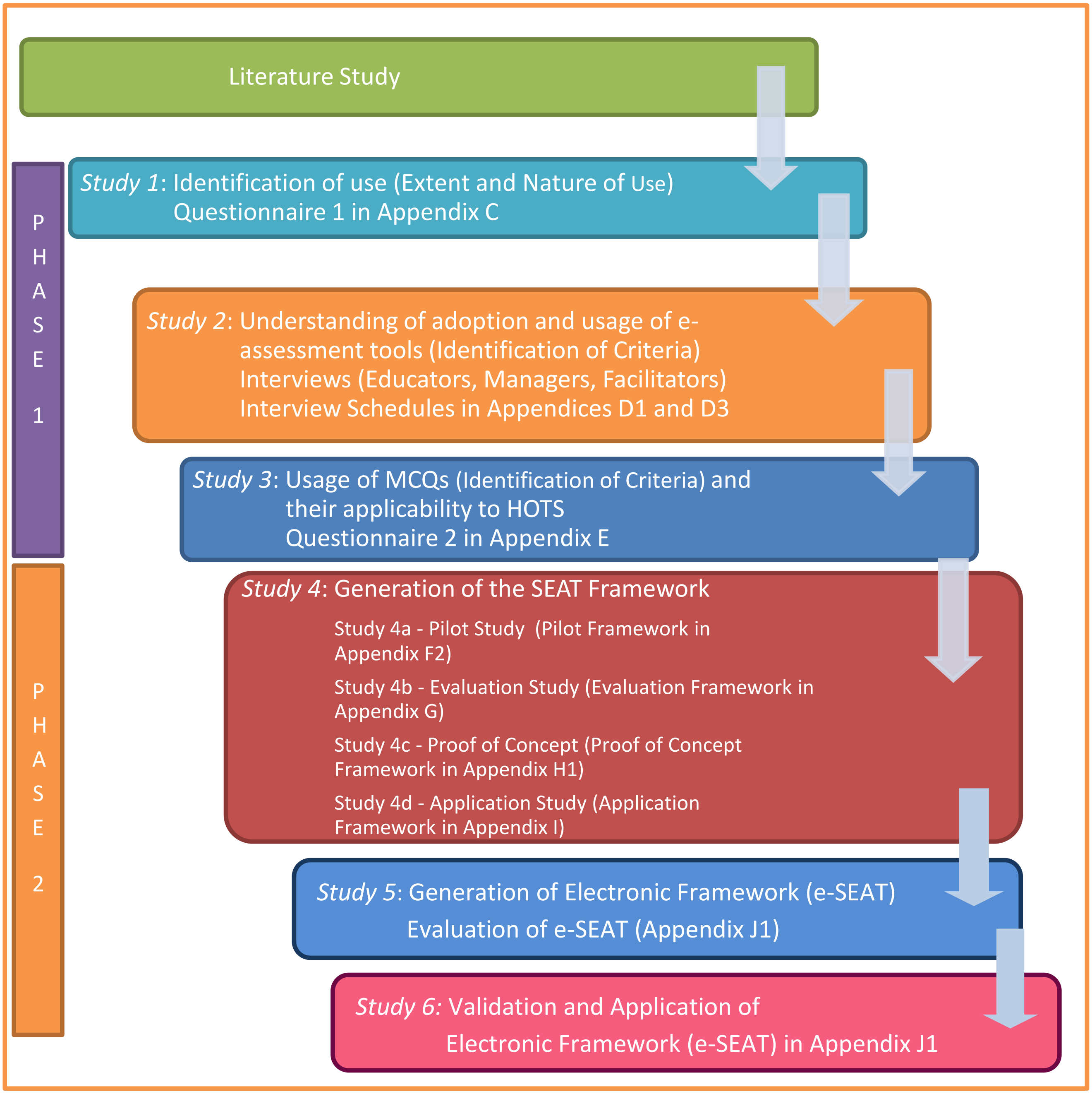


Figure 1.3: Outline of study

1.8.10 Ethical issues

As previously stated:

- Directors of schools, or heads of departments at the institutions where questionnaires were distributed, were contacted for Gatekeeper Consent prior to Studies 1 to 3.
- An ethical clearance application and the research instruments were presented to the Ethical Clearance Committee of the College of Science, Engineering and Technology at UNISA (see Appendices A1 to A3).
- An ethical clearance application and the research instrument for Questionnaire 1 were presented to the Ethical Clearance Committee at UKZN, for their approval (see Appendix B). UKZN did not require ethical clearance for the subsequent studies in this research.
- Anonymity of participants was ensured and their identities will not be revealed during the write-up of the results or in any publications emerging from this research.
- Participants in all surveys completed informed consent forms.

1.9 Structure of the thesis

Table 1.2: Outline of chapters in thesis

Chapter outline for thesis
1. Introduction and background
2. Literature Study: electronic assessment (e-assessment)
3. Literature Study: multiple choice questions (MCQs) in e-assessment
4. Research design and methodology
5. Data presentation and analysis of Phase 1 Studies
6. Data presentation and analysis of Phase 2 Studies
7. Conclusion and recommendations
References
Appendices (presented on a CD, attached to the back cover of this Thesis)

As depicted in Table 1.2, Chapter 1 introduces this study and outlines its background. The literature studies in Chapters 2 and 3 present terms, concepts, and attributes associated with assessment in general and e-assessment and MCQs in particular. International usage and practices in e-assessment, are briefly reviewed. In Chapter 3 an initial synthesis of evaluation criteria from the literature, is outlined from which to derive the evaluation framework presented as a result of this research. Chapter 4 sets out the research design and methodology to be adopted in this study. Chapters 5 and 6 discuss the results of the data collection of Study 1 to Study 6. Chapter 5 addresses Phase 1 of the action research process. Phase 1 includes Study 1 which identifies the 'extent and nature of use' of assessment tools in the context of South African computing education and Study 2 which uses interviews to expand Study 1. Study 3 presents the investigation into the types of MCQs adopted by South African academics, as well as the applicability of these to higher order thinking skills (HOTS).

The progression of the SEAT Framework from the Pilot Study, through to the Evaluation, Proof of Concept and Application Studies, is outlined in Studies 4a, 4b, 4c and 4d in Chapter 6, which sets out the work done in Phase 2 of the action research series. Chapter 6 also explains the creation of the prototype electronic version of the final e-SEAT Framework. In Study 5, the e-SEAT Framework is evaluated for both its criteria content and applicability. Study 6 validates the e-SEAT Framework through its application to existing e-assessment systems.

Chapter 7 concludes the study and summarises its findings. It revisits the research questions and reviews the process and contribution of this research. Recommendations are made and directions noted for future research. The References follow thereafter.

The Appendices are included on a CD, attached to the back cover of this Thesis.

1.10 Chapter summary

This chapter commenced with Section 1.1, an introduction to the study. Section 1.2 focused on the background to the study, including a general motivation for the research (Section 1.2.1) and the specific motivation (Section 1.2.2). The problem statement and research questions were presented in Sections 1.3 and 1.4 respectively. The benefits of the study (Section 1.5) were discussed under two subsections – namely the potential contribution of this study (Section 1.5.1) and the beneficiaries of the study (Section 1.5.2). A brief outline of the literature study was presented in Section 1.6. The scope of the study was discussed in Section 1.7 with Section 1.7.1 focusing on the domain and context, Section 1.7.2 presenting the delimiters and limitations, and Section 1.7.3 discussing the assumptions. Section 1.8 was a comprehensive discussion on the research design and methodology, discussing the approaches used in each of the substudies of this research. Section 1.9 outlined the structure of the thesis.

CHAPTER 2 Literature study: electronic assessment (e-assessment)

Following the background to the study presented in Chapter 1, this chapter discusses the literature reviewed, with particular reference to literature on assessment; its terminology, types, purpose and methods. Furthermore, it provides a foundation for the discussion in Chapter 3, which specifically describes the types of multiple choice questions and sets out categories outlined in the literature that are used to design or evaluate e-assessment systems.

This chapter starts by introducing the general literature on assessment, including discussions on its definition (Section 2.1), types (Section 2.2), purpose (Section 2.3), and measures (Section 2.4). Section 2.5 focuses more specifically on e-assessment, outlining its definitions (Section 2.5.1 and Section 2.5.2), features and components of e-assessment tools (Section 2.5.3 and Section 2.5.4 respectively); common e-assessment tools adopted (Section 2.5.5) and the procedures followed by e-assessment systems (Section 2.5.6); as well as the benefits (Section 2.5.7), disadvantages (Section 2.5.8), constraints (Section 2.5.9) and solutions to the constraints (Section 2.5.10) associated with its adoption. Section 2.6 concludes the content by presenting the chapter conclusion.

2.1 Definition of assessment

The term assessment is defined by different individuals or institutions in many ways, sometimes even with different goals. This section presents some general definitions of assessment in an attempt to fully understand the concept. Rovai (2000) describes assessment as an important, continuous phase of both teaching and learning, which supports the process of collecting, describing, or quantifying information about student performance. Anderson, Ball and Murphy (1975) add an additional facet to assessment, stating that it is usually complex since it focuses on various important outcomes which require a number of 'multisource/multijudge' techniques for evaluation.

McAlpine (2002) describes assessment as a means of communicating with various stakeholders, including students, educators, curriculum designers, administrators and employers, who each obtain some form of feedback from these assessments.

These stakeholders include, McAlpine (2002:4):

- 'students – on their learning,
- educators – on their teaching,
- curriculum designers – on the curriculum,
- administrators – on the use of resources, and
- employers – on the quality of job applicants'.

Therefore, assessment has a far-reaching effect in that it may affect decisions about grades, placement, instructional needs and curricula. In essence, as Souali, Afia, Faizi and Chiheb (2011) point out, assessment is a part of the learning process used to understand better the students' current knowledge through a process of identifying, gathering and interpreting data on their performances and progress.

Assessment is most effective when student confidence in the marker (assessor) is high, which reinforces the vital requirement to assess accurately and consistently, as well as the importance of providing useful and understandable feedback to the student (Jordan, 2011).

2.2 Types of assessment

In the traditional form of assessment, students of a single class are assessed using a common procedure at an officially controlled location or, if student numbers are large, simultaneously at various locations. Despite traditional assessment methods being reliable and consistent, current trends are focusing more on 'student-centred active learning' and assessments which commonly include some element(s) of electronic assessment. However, it is important to note that the assessment principles presented in Section 2.4 for direct student assessments in traditional learning environments, remain the same for online situations (Rovai, 2000).

Assessment can be classified in multiple ways; some classifications include diagnostic, formative or summative assessment (Mostert, de Bruyn, & Pretorius, 2012). Brief descriptions of these types are presented below:

2.2.1 Formative versus summative

Formative assessment is an ongoing measurement designed to assess students' knowledge and skills with the intention of supporting them in their ongoing learning experiences (Souali et al., 2011). Formative assessment includes self-assessment and diagnostic assessment, and focuses on providing feedback to students to highlight areas for further study with the goal of improving individual future performance (McAlpine, 2002). Since in formative assessment, students are expected to learn from the questions they got wrong, feedback is both necessary and essential (Alton, 2009; Pattinson, 2004). The format of the feedback from the educator can either be written or oral (Souali et al., 2011).

When feedback is used correctly and extensively, the process is thus bidirectional between the educator and the student. Feedback serves a dual purpose – it provides the educator with a deeper understanding of individual student abilities and also supports students in improving their performance through enhancing, recognising and responding to the students' understanding of material presented. Karl, Graef, Eitner, Wichmann, Holst and Beck (2011) explain that when students personally monitor their academic progress, it helps to promote independent learning which can lead to sound learning strategies, better acquisition of skills, more effective study processes, and higher achievement.

For formative assessment to be effective, the feedback provided must be useful. Souali et al. (2011), refer to three forms of feedback; namely, feedback:

- regarding the result,
- about the students' mistakes, and
- about how to proceed next.

The aim of formative assessment is to support learning. Thus it is referred to as 'assessment *for* learning' (Mostert et al, 2012). Formative assessment also provides students with an opportunity to engage with learning material so they can prepare for summative assessment (Dube & Ma, 2011).
In the South African university situation, formative assessment refers to those tests, projects and revision assignments during the year or semester that do not contribute to the year mark, hence they would usually be adopted for self-assessments. Locally, much of the semester or year work does,

however, contribute to the final mark; therefore it comprises part of summative assessment, which is discussed next.

The effectiveness of e-assessment can be measured during a learning programme after completion of a learning activity (Khedo, 2005; McAlpine, 2002; Pretorius, Mostert & de Bruyn, 2007). e-Assessment facilitates timeous feedback, often while students are still focused on the learning material (Cook & Jenkins, 2010). McAlpine (2002) presents useful definitions of forms of e-assessment, several of which are addressed in the next few sections.

Summative assessment is a quantitative measure, usually given at the end of a course to evaluate the progress and development of a student at a particular time (Cook & Jenkins, 2010; Souali et al., 2011). The main focus of summative assessment for the educator is to obtain information about students' performance; thus it is referred to as 'assessment *of* learning' (Mostert et al, 2012). As opposed to the term used in the previous paragraph, 'assessment *for* learning', when referring to formative assessment, 'assessment *of* learning' usually means monitoring students' performance against the objectives to assist the educator in determining ways to improve future teaching and learning processes. Since answers to summative questions do not require feedback, some visual indication is however necessary to indicate what was correct and what was incorrect, either after each question, or at the end of the test, as part of a review (Alton, 2009). (Researcher's note: In the case of a formal examination, however, the final mark is often the only feedback and students do not see the marked examination script).

Summative assessments are not designed to give immediate or continuous feedback, but rather to give an indication of what has been learned up to that point (Souali et al., 2011). Thus, the results of summative assessments – which are designed to judge the students' overall performance – are also useful for external parties, such as prospective employers, who might base their decisions on the information gathered from summative assessments (McAlpine, 2002). Concise summaries of students' abilities are available from summative assessments and should be easy to interpret. The role of the educator in summative assessments is that of an adjudicator who judges a student's level of achievement at a particular point in time (Souali et al., 2011). In the South African university context, summative assessment also includes those tests and projects during the year or semester, which contribute to the year mark for the module (Dube & Ma, 2011). Assessed tests and projects are, however, returned to students as feedback.

While e-assessment is well suited to formative assessment, its use in summative assessment is limited, since the ‘high stakes for the subject taught requires different skills to those which can be assessed via e-assessment’ (Miller, 2012: 1). Certain learning content can only be assessed effectively by means of written answers.

Computer-assisted assessment can be used for summative assessment, along with assessment by marked coursework, with feedback presented to students. This form of assessment is formal, structured, and invigilated just as in a traditional paper-based examination. These assessments can be done at different times in the module including at the end, or at predetermined times during the course, to determine a value which forms a final mark reflecting the student’s performance. This form of assessment can serve as an extrinsic motivator for students (Kadhi, 2004; Khedo, 2005; Mc Alpine, 2002).

2.2.2 Formal versus informal

Formal assessments adopted for summative rather than for formative purposes, are typically assessments where students are aware that the task being undertaken is for official assessment purposes. Research indicates that some students view this type of assessment as fairer, more explicit and less biased; while other students feel pressurised by such assessments and may learn facts superficially and perform well, yet without a deep understanding of the material (McAlpine, 2002).

Informal assessments are best used for formative or diagnostic tasks, since they are not data-driven but rather content-driven and performance-driven assessments. These assessments provide the educator with unique information that helps promote student-centred learning. In particular, educators can gather the unique behaviour of a student that will add value to the delivery method and technique they adopt (Banks, 2012). This can help to reduce students’ anxiety associated with formal assessments as it presents a deeper understanding of a student’s abilities, due to their formative nature (McAlpine, 2002).

2.2.3 Final versus continuous

Final assessment takes place at the end of a module or course. It can be simple to organise and less time-consuming as the assessment is condensed into a short time duration. It is most appropriate where each part of a field of study contributes to grasping of other sections (Falchikov, 2013). In such situations final assessment is conducted as a complete whole, rather than constituent parts separately. According to McAlpine (2002), final assessment cannot be used for formative purposes. However final assessment can be viewed as formative assessment when the student uses it as a foundation for the work in successive courses, for example, a project in first year can be a support to learning in the second year of study.

Continuous assessment takes place at intervals during a module or course. It often takes the form of coursework, combined with the final assessment (the examination) (Falchikov, 2013). The purpose is to provide both students and educators with feedback regarding performance in a test or in other deliverables. The final result from continuous assessment is based on evidence gathered over the duration of the learning period (McAlpine, 2002). Although the workload of the educator is increased in this form of assessment, the information provided to the educator can help to improve teaching and learning. Thus, continuous assessment is most appropriate when a student's capabilities are assessed over a series of pieces of information (McAlpine, 2002). e-Assessment is a powerful tool to enhance the use of continuous assessment for providing rapid and detailed feedback to both students and educators about the learning process.

2.2.4 Process versus product

Process-driven assessments assess students' skills or abilities in the context of a particular task, while *product-driven assessments* are appropriate where knowledge content is fundamental, as these assessments can be easily summarised and generally have more tangible criteria, which makes them easier to create (McAlpine, 2002).

2.2.5 Convergent versus divergent

Convergent assessments have a single correct answer that is required from the student, and thus are easier to mark both by automated and manual systems, but are well suited to e-assessment. Delivery and feedback are usually faster on this type of assessment, due to the nature of answers required. They can cover wider curriculum content since they are so specific in nature. However, they can often be limited in scope. Convergent assessments assist the educator to discover *whether* a student knows, understands or can perform a predetermined task (Swaffield, 2008). A disadvantage with convergent assessments is that educators are often tempted to test only the concepts that can be easily translated easily into convergent form. This may result in poor assessment quality. As stated, e-assessment is best suited to convergent assessments; however, the questions and tests need to be skilfully designed (McAlpine, 2002).

Divergent assessments allow a range of possible answers from the student, based on his/her understanding and knowledge. They are more authentic and have the potential to test higher cognitive skills, often termed higher order thinking skills (HOTS). Divergent assessments assist the educator to discover *what* the student knows, understands or can perform (Swaffield, 2008). They can be time-consuming to set and mark, and require greater marking skill than convergent assessments. Thus the human assessor should be well trained, or provided with detailed marking criteria (McAlpine, 2002). In the case of e-assessments where more than one answer is right, the program must be able to recognise all the correct options.

This section has overviewed various ways of categorising assessment and is relevant to both conventional assessment and marking and also to e-assessment.

Despite the multiple methods of classification available for assessments, e-assessment has become accepted as a key tool for evaluating student performance through a range of these types, including diagnostic, formative and summative assessment (Belton & Kleeman, 2001).

2.3 Purpose of assessment

As expressed by Horton and Horton (2003: 288), assessment usually aims to quantify the effectiveness of learning, ' …. it is seldom an end in itself, but rather, an important element of all courses'.

Lambert and Lines (2013) support this argument by stating that assessment is sometimes perceived as a 'necessary evil', which serves more to support educators than students. In reality, however, assessment can help to explore different ways of thinking about the subject matter being taught.

The purpose of any given assessment can be ascertained by identifying:

- the reason assessment is being conducted,
- how best the assessment should be designed to meet particular requirements,
- what decisions can be made from the assessment results,
- what information must be gathered to make these decisions, and
- the methods most effective for gathering the required information (McAlpine, 2002).

This is supported by the ten advantages of testing outlined by Roediger, Agarwal, McDaniel and McDermott (2011). Testing:

- has a cognitive effect: retrieving information from memory supports subsequent retention,
- identifies gaps in knowledge,
- helps students to learn more from the next study episode,
- produces better mental organisation of knowledge,
- improves transfer of knowledge to new contexts,
- can facilitate retrieval of material that was not tested,
- improves meta-cognitive monitoring,
- prevents interference from prior material when learning new material,
- provides feedback to instructors, and
- frequently encourages students to study.

2.4 Measures of assessment

According to McAlpine (2002), Ashcroft and Palacio (1996), and Falchikov (2013), assessment is of value only if meaningful measures of comparison exist, including:

- validity, which means that the assessment tests a relevant skill or ability,
- reliability, which is obtained if a student will achieve the same result in a repeated assessment,
- referencing, which indicates that an assessment is meaningful if the student's abilities can be compared with a common measure such as other students' performance, objective criteria identified by the educator, or the student's own performance in another area,
- quality, aiming for an assessment to be set at approximately the difficulty level of the average student, but should also differentiate between the students to allow the educator to separate the students as much as possible, based on their understanding of the material tested, and
- grades (in South Africa, referred to as 'marks') awarded to students, should be easily understandable by the student or any external party, as they represent concise summaries of students' performances.

2.5 Electronic assessment (e-assessment)

Current computing and electronic technology offer ways of enriching educational assessment both in the classroom and in large-scale testing situations. As the digital divide decreases, educational technology should be applied in ways that capitalise on these new frontiers of innovative assessment, generating rich new assessment tasks and effective scoring, reporting and real-time feedback mechanisms, for use by both educators and students (Scalise & Gifford, 2006).

With student numbers increasing, and universities' funding decreasing, e-learning is seen as a potential solution to the issue of quality in assessments (Govender, 2003), ' … in the present context of financial stringency and greatly increased numbers of students entering Higher Education, the maintenance of quality in the face of reduced units of resource is something to which technology-assisted teaching might contribute' (Catley, 2004: 1). Quality assessments and quality assurance mechanisms should be integral aspects of university procedures (Bull, 1993).

The use of automated assessment can assist in providing detailed, individualised and instant feedback to large numbers of students, through the use of a mastery learning model, in which students can repeat or progress at their own pace. Such mastery learning and automated drills for practice (Alessi & Trollip, 2001; Clariana, Ross & Morrison, 1991) are ideally implemented by e-assessment in the form of multiple choice questions (MCQs). However, MCQs can also be used creatively with a range of other approaches and methods, for example, for peer-assessment and for self-assessment, short answer tests, closed tests, and Information Technology (IT) projects (Luckett & Sutherland, 2000).

Besides providing feedback, e-assessment provides an attractive option for higher-education institutions facing the logistical problems associated with the increase in student numbers (Bull & McKenna, 2003). Furthermore Bull and McKenna point out that the consistency of electronic marking removes concerns associated with subjective manual marking by the human assessor.

Learning benefits provided by e-assessments are derived when students reinforce their understanding of core concepts through repetition of material, or by taking a variety of assessments on the subject matter. Furthermore, through the provision of timely feedback that indicates their mistakes, students are able to close the gap between actual and desired performance levels (Nicol, 2007; Walker, Topping & Rodrigues, 2008).

2.5.1 Definition of e-learning

e-Learning, web-based learning (WBL), and online learning are terms often used interchangeably; yet, according to Tsai and Machado (2002), they represent concepts with subtle, yet important differences. Turban, King, Lee, Liang and Turban (2010: 68) supports this view, stating that 'e-Learning is broader than the term online learning, which generally refers to purely web-based learning'. Tsai and Machado (2002) further describe web-based learning as learning materials delivered in a Web browser; however, their definition includes materials packaged on CD-ROM or other media. They refer to content readily accessible on a computer via the Web or the Internet, or simply installed on a CD-ROM or the computer hard disk, as online learning, not e-learning.

Clark (2003) defines e-learning broadly as instruction provided through technology, more commonly through the use of a CD-ROM, the Internet, or an Intranet, which:

- presents content relevant to the learning objective,
- uses interactive instructional methods, such as examples and practice, to facilitate learning,
- adopts multimedia elements such as voice, pictures or moving images to deliver content, and
- builds new knowledge and skills linked to individual learning objectives.

In a more current publication, Turban et al. (2010) extend the concept of e-learning to include online delivery of information, not only for formal education, but also for training or general knowledge management. Such systems are usually web-based, making knowledge accessible to those who need it, when they need it, anywhere, anytime. However, formats can vary, ranging from virtual classrooms through to mobile learning (m-learning) applications, by which material is delivered wirelessly to students via mobile phones or Personal Digital Assistants (PDAs) (Turban et al., 2010).

In this study, an all-encompassing definition of e-learning is adopted, which incorporates a broad range of educational technologies and types of learning/instruction. e-Learning is viewed as including interactive institution-wide learning management systems, web-based teaching materials and hypermedia, multimedia CD-ROMs, e-learning tutorials, simulations, games, and e-assessment (Alessi & Trollip, 2001; de Villiers, 2012a). In the realms of Web 2.0 and e-Learning 2.0, where students are not only consumers of content, but also contributors, there is a major role for collaborative software technologies on the Internet such as discussion boards, e-mail, blogs, wikis, chat rooms, academic use of social networking sites, and educational animations (Ebner, 2007; Turban et al., 2010).

2.5.2 Definition of e-assessment

According to Dube and Ma (2011), e-assessment is one of the domains of e-learning where Information and Communication Technology (ICT) is used to present assessments and record students' responses. Cook and Jenkins (2010), refer to e-assessment as assessment that is stored, delivered, answered and often fully marked automatically, using some form of technology. Similarly, Byrnes and Ellis (2006) point out that so-called computer-based assessment (CBA) is considered to be a rapid and accurate tool for the assessment of students' learning. Significant developments in this area have resulted in it being increasingly implemented for student evaluation worldwide. Cook and Jenkins (2010) state that e-assessment is distinctly different from computer-assisted assessment. The JISC (2007) define e-assessment as a range of activities – such as designing and delivery, marking and processes of reporting, storing and transferring of data, of assessments – in which digital technologies are adopted.

Although e-assessment (Costagliola & Fuccella, 2009; Sangi, 2008) is now the most common term used for online assessment or automated assessment methods adopted in both e-learning and traditional class-based learning, there are synonyms such as:

- Computer-aided testing (CAT) (Karl et al., 2011),
- Computer-administered tests (Waring, Farthing & Kidder-Ashley, 1999),
- Computer-aided assessment (CAA) (Brown, Bull & Race, 2013; Costagliola & Fuccella, 2009; Davies, 2001; Duarte, Nunes, Neto & Chambel, 2006; Falchikov, 2013; Fielding & Bingham, 2003; Lambert, 2004; Sim, Holifield & Brown 2004; Weerakoon, 2001),
- Computer-assisted assessment (Brown, Bull & Race, 2013; Conole & Warburton, 2005; Costagliola & Fuccella, 2009; Govender, 2003; Khedo, 2005; Mostert et al, 2012),
- Computer-based assessment (CBA) (Byrnes & Ellis, 2006; Costagliola & Fuccella, 2009; Harper, 2003; Khedo, 2005; Miller, 2012),
- Computer-based testing (CBT) (Govender, 2003; Jordan, 2011; Mostert et al, 2012; Tsintsifas, 2002),
- Computerised tests (Alessi & Trollip, 2001),
- Electronic assessment (e-assessment) (Bialo & Sivin-Kachala, 2013; Brown, Bull & Race, 2013; Dube & Ma, 2011; Engelbrecht & Harding, 2003; Govender, 2003; Jordan, 2011; Jordan, 2013; Mostert et al, 2012; Patel, Kothari & Makwana, 2013; Rusman, Boon, Martinez-Mones, Rodrigues-Triana & Retalis, 2013; Tsintsifas, 2002),
- Interactive computer marked assessment (Jordan, 2011),
- Online Assessment (Besterfield-Sacre & Shuman, 2008; Govender, 2003),
- Online evaluation (Nelson, 1998),
- Online examinations (Khare & Lam, 2008),
- Online testing (Costagliola & Fuccella, 2009; Fielding & Bingham, 2003; Horton & Horton, 2003; Lambert, 2004),

- Technology Mediated Assessment (Besterfield-Sacre & Shuman, 2008), and
- Technology-enhanced assessment (Jordan, 2011).

The term of preference in the present study is e-assessment. The focus of the work in the present study, however, as stated in the Disclaimer in Section 1.1, is the adoption of e-assessment of the MCQ genre for both summative and formative assessments. It excludes aspects of e-assessment such as e-portfolios, blogs, wikis, peer assessment, etc.

2.5.3 Features of e-assessment tools

According to Khedo (2005: 188), e-assessment presents 'a new way of harnessing the power of computers to the field of education'. Use of this potential in assessment supports both educators and students in:

- the method of delivering assignments and examination papers that are appropriate for automated assessment,
- setting up marking memorandums and analytical tools for diagnosing and correcting the work submitted by students, and
- generating automated reports and consolidating students' results (Khedo, 2005).

e-Assessment tools are most commonly adopted in situations of increasing class sizes and the associated demands on educator time and resources (Sim, Holifield & Brown 2004). When the questions are carefully designed, these tools can also be adopted to measure knowledge, comprehension and application of learning outcomes (Souali et al., 2011). The issue of assessing higher order thinking skills (HOTS) is addressed in Sections 2.5.8.4 and 3.1.3.

Features such as detailed feedback and the ability to repeat a test are present in most e-assessment systems (Maurice & Day, 2004). Prompt feedback is usually a characteristic of e-assessment. Rapid feedback can help to guide students about the educator's expectations at an early stage of learning (Khedo, 2005). e-Assessment systems allow educators to assess the students both with formative and summative objectives (Costagliola & Fuccella, 2009), concepts that have been addressed in Section 2.2.1. Formative assessments (Section 2.2.1) are administered during the learning process to give information on the learning state of each student, and thus allow the student to pay more attention to areas of the curriculum that prove problematic, or the educator to take remedial action as required; while summative assessment (Section 2.2.1) occurs mainly at the end of the learning process and is used to express a judgement of the learning state of each student. Karl et al. (2011) demonstrated the positive impact that formative e-assessment has on first-year students' learning and performance.

The emergence of e-assessment, particularly for self-assessment tests – with integrated multimedia learning material, gives educators the opportunity to adopt testing procedures in computer laboratories, presenting students with sets of questions that are usually in the form of multiple choice questions (Ventouras et al., 2010). Successful adoption of e-assessment systems encourages the educator to focus on the actual assessment process, rather than the potential of the technology alone. Systems that do not align the actual assessment process with learning objectives result in an assessment 'backwash', where students tend to learn only what they think they will be examined on. Therefore, with e-assessment, as in conventional assessment, the actual assessment should be objective, criterion-based, reliable and valid (Ventouras et al., 2010).

Although e-assessment tools were initially designed primarily to assist educators in grading tasks, referred to in South Africa as marking, these tools have evolved to become valuable ways of fostering self-directed learning, especially if they are freely available for students to use in their own time, getting support and assistance from the automated feedback facilities (Sim, Holifield & Brown, 2004). Many higher education institutions are currently adopting e-assessment tools either for grading students or for providing feedback to enhance the learning experience (Sim, Holifield & Brown, 2004). Miller (2012) suggests that e-assessments are valuable in self-assessments, as students spend time increasing their knowledge and redoing formative assessments until they get a good mark. e-Assessment, through the use of self-assessments, forces students to engage more with the subject material, thus helping them to focus more in the class and also encouraging them to read the textbook. This results in them becoming motivated and ultimately achieving better marks in summative assessments (Miller, 2012).

2.5.4 Components in e-assessment tools

Whatever the terminology used, e-assessment usually involves presenting the student with a variety of questions online, after which they respond online, and finally their responses are marked electronically.

The results received from the software are sent to a database. This allows the results to be stored for later use so they can be accessed by educators. At the discretion of the educator the results may be presented to the student immediately (Maurice & Day 2004).

e-Assessments may be stand-alone and specific to certain machines within a computer laboratory, or based on a local network (intranet) or, as is increasingly common, web-based. They can be either supervised or non-supervised, with the option of allowing students to check their own progress through self-assessment (Khedo, 2005).

The main paradigm behind web-based software is that these applications are usually platform independent hence they can run on any operating system, and therefore do not necessitate software installation (Khedo, 2005). In some cases, an e-assessment is presented via an external medium such as a CD; but in such cases the results cannot be transferred to a database or web environment.

The databases in an e-assessment system typically store the following items:

- the questions,
- the answers,
- the composite tests – which incorporates the questions and the answers,
- details about users (students and administrators),
- the students' responses, and
- the students' results.

The database is central to the e-assessment system, as it aids the management of the assessment data, as well as the flow of the data through the system. It stores all the available questions, their answers (for automatic marking), comparison and classification by question types, mark value of questions, data variables for students' details, answers given by students, as well as overall record, grades and scores of students (Dube & Ma, 2011).

Some e-assessment software allows educators to create a bank of questions which can be stored in their database. This facilitates automatic random generation of test items by the e-assessment system, based on the predefined parameters set by the educator. These parameters may include the number of questions from a particular section of the learning material, and/or the number of questions from each level of difficulty.

Although MCQs in e-assessment are more commonly used for testing lower-order skills (such as knowledge, understanding and application), if these questions are designed properly they can also be used for testing higher-order skills (such as analysis, synthesis and evaluation). This will be addressed in Sections 2.5.8.4 and 3.1.3. However, due to the nature of the questions adopted in e-assessment, the tool adopted facilitates the automation of what was previously a very time-consuming task – marking scripts and monitoring the progress of students (Khedo, 2005).

2.5.5 Common e-assessment tools

e-Assessment tools mentioned in the literature (Fielding & Bingham, 2003; Horton & Horton, 2003; Lambert, 2004; Miller, 2012 ; Pretorius et al., 2007; Seale, 2002; Shulman, 2005; Singh & de Villiers, 2010; Tsintsifas, 2002) include:

- Assignment in Moodle,
- Blackboard Learn by Blackboard,
- Blackboard Analytics by Blackboard,
- Coursebuilder for Dreamweaver by Macromedia,
- CourseCompass by Pearson,
- ExamView Test Generator by Pearson,
- EzTest Online by McGraw Hill,
- HostedTest by HostedTest.com,
- Hot Potatoes by Half-Baked Software,
- MarkIt
- Perception by Questionmark,
- Quick Rocket by LearningWare,
- Quiz in Moodle,

- Random Test Generator Pro by Hirtle Software,
- SAM by Cengage
- Tests and Quizzes in Sakai,
- Test Generator by Fain & Company,
- Umfundi by FullMarks,
- Unit-Exam by Unit-Exam.com, and
- Virtual Assessor.

Earlier e-assessment systems such as Question Mark for DOS, and its successor Question Mark for Windows, were relatively simple packages that enabled their users to create and to run questions on a computer. Ideally, they were designed to offer practical solutions for practice tests, self-assessment and limited formative testing. However, due to the lack of industry standards at the time of their development, they did not comply with any (Khedo, 2005). Furthermore, in the early stages, there were few guidelines on what comprised a good e-assessment package. Yet, even at the early stage of development, it was envisaged that computerised testing would evolve to become a very useful tool. e-Assessment has now become an essential and integral part of many e-learning packages.

e-Assessment tools can also be classified as 'expression-oriented' – where the educator specifies a single correct solution which is compared verbatim with a student's response through an equivalence test, or 'property-oriented' – where the educator defines an expression specifying the properties or conditions the student's response must satisfy in which case, there is more than one correct answer (Duarte et al., 2006).

2.5.6 e-Assessment procedures

Irrespective of the tool selected, the deployment of e-assessment tools follows a similar cycle of developing, conducting and reporting assessments. This section presents some e-assessment cycles.

2.5.6.1 Horton's e-assessment cycle

A six step cycle is depicted in Figure 2.1.

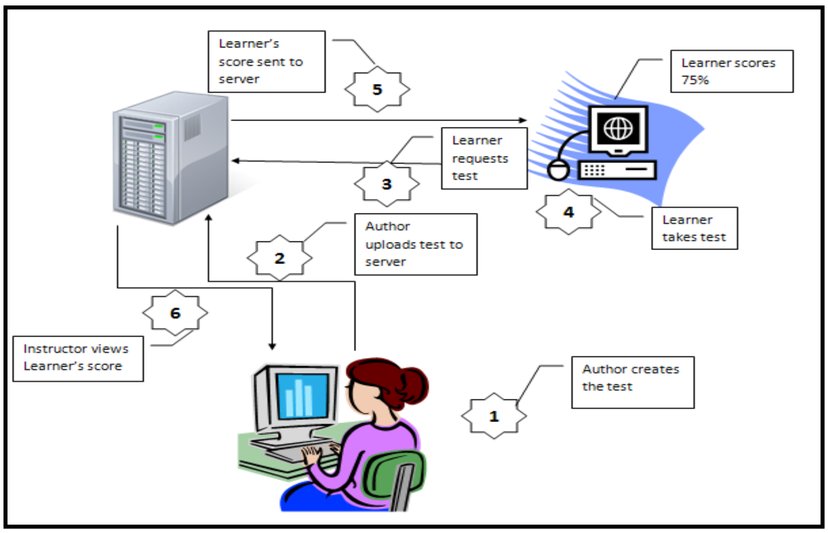


Figure 2.1: The cycle of developing, conducting and reporting tests (Adapted from Horton & Horton, 2003: 328)

The process suggested by Horton involves the following:

- initiation of the test by the author, who creates the test with an assessment tool (1 in Figure 2.1),
- uploading of the e-assessment to the server (2) where it can be accessed by students at a specified date, time and locality (3, 4),
- displaying and storing of the students' results after the assessment is completed (5), and
- monitoring the progress of the students by the assessor (6) (Horton &Horton, 2003).

2.5.6.2 Khedo's e-assessment cycle

Similarly, Khedo (2005) outlines five processes usually followed in e-assessment systems:

- presenting the scenario or question to the student,
- obtaining a response entered by the student on the computer,
- evaluating the student's response,
- providing a mark or score to the student, and
- providing feedback to the student.

All the above processes require that the underlying system be reliable.

2.5.6.3 Dube and Ma's e-assessment cycle

Dube and Ma (2011) compress two of the above stages into one and produce four in all:

- presentation of the assessment activity,
- recording of student responses,
- automatic assessment of student responses, and
- presentation of feedback.

Generally, any e-assessment system includes the following procedures:

- access to the system by running the executable file or via the web link,
- authentication of students through the use of log-in details which are also used to map feedback to each individual student,
- presentation of the assessment, both the instructions and the questions,
- responses entered by the student to the questions presented,
- recording of student's responses to the questions presented,
- marking of the student's responses, automatically, and
- feedback presented to students either in the form of marks gained, correct responses to questions highlighted, or sending of feedback to the individual student in a personalised manner (Dube & Ma, 2011).

2.5.7 Benefits of e-assessment

e-Assessment holds certain advantages over manual assessment. Mostert, de Bruyn and Pretorius (2012) summarise these benefits by indicating that a variety of question types can be adopted, questions may include multimedia elements, immediate feedback can be provided to students, statistical analysis gleaned from the system can help educators to improve test items, and feedback can be given to educators regarding gaps in students' understanding; and self-assessment opportunities are available to students.

Various benefits associated with the usage of e-assessments, as gleaned from the literature, are discussed below, categorised under the subheadings of productivity, reduced cheating due to randomisation, subjectivity and bias, efficiency, enhanced reporting and statistical analysis, feedback to students, time/space independence, more frequent testing without increased marking overhead, question banks and formats, cost savings, and holistic evaluation of courses.

2.5.7.1 Productivity

Academics, globally, face an increasingly pressured work environment due to the rise in student numbers at universities, coupled with a decline in resources. High student-to-educator ratios create a heavy load for educators and increase the effort of keeping track of how students progress. The evaluation process becomes more difficult and the grading processes become exhaustive and time-consuming. This has led to universities implementing new techniques of assessing students, and, in particular, to an increase in the usage of e-assessment (Akinsanmi, Agbaji, & Soroyewun, 2010; Bani-Ahmad & Audeh, 2010; Byrnes & Ellis, 2006; Sim, Holifield & Brown, 2004). e-Assessment is suited for both high and low-volume needs, since automated assessment harnesses the power of technology to deliver assessment via the web or an intranet (Dempster, 2007; Tsintsifas, 2002).

e-Assessment allows educators to administer assessments regularly, especially with large student numbers, without the additional burden associated with multiple manual assessments per student. This process is fairly effortless if large databanks are available, since marking is done automatically (Mostert et al, 2012). e-Assessment tools are thus both time and cost effective to improve learning (Miller, 2012).

2.5.7.2 Reduction of cheating due to randomisation

A direct implication of working in computer laboratories is that space is limited for students sitting for evaluation examinations. Space is vital to reduce cheating (Schuwirth, Vleuten & Donkers, 1996). Byrnes and Ellis (2006) established that with randomisation functionality in the presentation of e-assessments, cheating was reduced. The incidence of cheating was also found to decrease through the randomisation of both questions and distracters (Schuwirth, et al., 1996). e-Assessment may permit switching answers by allowing the student to devote more 'time-on-task', thus also helping to discourage cheating (Byrnes & Ellis, 2006). However, to ensure that the assessment is fair to all students, the educator must have confidence in the uniformity of his/her questions in situations when they give different questions to different students, yet are assessing the same concepts (Miller, 2012).

2.5.7.3 Subjectivity and bias

The 'large class' problem is further exacerbated by subjectivity on the part of educators, due to:

- exhaustion, since the standard of manual grading declines considerably as the educator marks more answer sheets,
- subjective bias by personal contact between the educator and the student and the educators' assumed knowledge of each student's abilities,
- the 'middle-mark bunching' syndrome, which occurs when examiners tire and are comfortable to allocate a middle-range mark to most students, depriving them of the actual mark deserved by their answer, and
- the personal characteristics of a hand-written answer sheet, such as handwriting quality, style, and layout of scripts

(Bani-Ahmad & Audeh, 2010; Byrnes & Ellis, 2006; Simkin & Keuchler, 2005).

e-Assessment avoids the problem of the above situations by focusing only on the correctness of the answer provided by the student. Similarly, Cook and Jenkins (2010: 3) describe computer marking as 'truly objective'. Byrnes and Ellis (2006) state it is a fair method to ensure that the assigned score reflects the students' true capabilities, and is not influenced by the assumptions or perceptions of the educator or by individual characteristics of the students. They further state that e-assessment helps to assess students' comprehension, critical thinking, and reasoning, rather than their memory or repetition of knowledge (Byrnes & Ellis, 2006).

Electronic marking helps to reduce human error, subjectivity and marker fatigue or subjectivity (Lambert, 2004). In addition, with automated marking, students need not be afraid of human criticism that is subjective in nature (Bani-Ahmad & Audeh, 2010). Thus e-assessment offers accurate, objective, and unbiased student evaluation (Byrnes & Ellis, 2006). Furthermore the 'halo' effect does not occur. With the halo approach, examiners are influenced by a well-written response early in the assessment and then grade the same student's next answers favourably, passing over the weaknesses, and vice versa (Byrnes & Ellis, 2006).

2.5.7.4 Efficiency

With the adoption of e-assessment, marking (grading) of assessments for large numbers of students is more efficient when done by automated means (Horton, 2000; Khedo, 2005; Maurice & Day 2004). Since assessments can be marked instantly, this can result in a great saving of educators' time in the longer term (Cook & Jenkins, 2010). These applications seek to reduce the strain that is placed on examiners with large classes, as tests are marked and recorded automatically by the software (Akinsanmi et al., 2010; Souali et al., 2011). Automated marking facilitates quicker, more detailed, and more accurate feedback, which is of benefit to both student and educator. e-Assessment provides an easy, fast and more manageable way of conducting tests for an increasingly growing population of students (Akinsanmi et al., 2010). It provides an improved and efficient assessment process by automated delivery, management, storage, and scoring of assessments (Byrnes & Ellis, 2006). Educators are enabled to assess a large number of students consistently within short time frames (Souali et al., 2011).

Furthermore, web-based e-assessment systems can provide '24-7' flexibility internationally, offering global access and anytime, anywhere usage through the medium of the Internet (Dube & Ma, 2011).

2.5.7.5 Enhanced reporting and statistical analysis

e-Assessment offers immediate scoring and reporting of students' test results, and simultaneously test security in a formal environment with controlled access. It can offer an exact duration for all test takers, without delays due to handling paper (Byrnes & Ellis, 2006).

Besides offering high quality content with large coverage, e-assessment can provide efficient and inexpensive data collection as well as procedures that help educators to improve their assessments (Byrnes & Ellis, 2006; Khedo, 2005). It can help educators to determine which questions seem to pose a problem for the students. Educators can review the problematic questions and make the necessary modifications/improvements in future assessment tasks (Byrnes & Ellis, 2006). Assessment results are typically stored in a format that enhances the process of producing statistical analysis of students' performance. Furthermore, e-assessments provide access to useful information about the students undertaking the assessments:

- which students attempt the e-assessment,
- how often they attempt the assessment, i.e. information regarding the retaking of formative assessments,
- how long it takes them to complete the assessment,
- what marks they achieved,
- the time at which they undertook the assessment, as well as
- cohort information that identifies the questions that students found easy and those that proved difficult. With this type of information, educators can investigate whether the problem was due to the way in which the question was phrased or whether the fact that a number of students got it wrong, suggests that the pertinent point should be addressed again in subsequent instruction (Catley, 2004).

e-Assessment can facilitate a smooth, automatic transmission and administration of marks between the university's information management systems and the databases which hold student records (Akinsanmi et al., 2010; Souali et al., 2011). The resulting diagnostic reports and analyses support educators in taking corrective or remedial action earlier as they identify areas of the curriculum that are not well understood by the students, or address the issue of how best to rephrase or present questions that are misunderstood.

The statistics provided by these software tools are easier to generate and analyse than the statistics provided by manual tests. Incorrect responses are usually clustered as percentages, thus allowing academics to easily determine which of the incorrect responses students most commonly selected (Souali et al., 2011).

The detailed reporting facilities of e-assessment tools can thus provide educators with valuable information regarding the misunderstandings of students, as well as helping students gauge their own level of knowledge, in cases where the consolidated records are made available to them (Mostert et al, 2012). Educators benefit greatly from the vast amount of valuable information provided by these statistical reports, on the performance both of students, as well as the actual examination questions set (Malau-Aduli, Assenheimer, Choi-Lindberg & Zimitat, 2013).

e-Assessment allows the educator to improve the assessment validity as inferred from the statistics of the assessment. By viewing statistics from subsequent, refined assessments, educators can ensure that the revised assessments fulfil their purpose. Students' progress can be monitored better when they are assessed more frequently. e-Assessment also allows the educator to test a wide range of topics within a body of knowledge. Both these features of e-assessment allow the educators to assess their students better, without increased marking commitments. Statistical analyses of test items and reports on student performance provide educators with feedback regarding gaps in student understanding of the learning material assessed. This, in turn, can assist educators in improving and enhancing their teaching practice.

2.5.7.6 Feedback to students

The key feature of e-assessment software is that it can provide students with detailed, constructive and consistent feedback, in a simple and efficient manner. Students complete an assessment and submit their answers, which are instantaneously marked and returned (Cook & Jenkins, 2010). Moreover, feedback on performances can be delivered instantly (Govender, 2003). For even the largest class sizes, marking is automated; thus reducing the workload of educators (Catley, 2004; Souali et al., 2011).

Functionalities for the examiner to set his/her questions online and for the students to take the test(s) and immediately receive automated feedback of their test scores, are features appreciated by academics internationally (Akinsanmi et al., 2010). Instantaneous and rich feedback provided to students on the adequacy of their preparation and knowledge is much more effective than feedback received after a number of days or even after several weeks. It fosters academic performance, allows students to identify key focus areas to be studied, and encourages independent learning, which in turn leads to self-efficacy (Duarte et al., 2006; Karl et al., 2011; Sim, Holifield & Brown, 2004). A majority of students in the study conducted by Karl et al. (2011) reported that the use of e-assessments enabled them to identify their personal strengths and weaknesses regarding their knowledge of course material. This also helped them prepare for examinations. These findings indicate increased student interest and motivation regarding e-assessment.

Furthermore, adaptive testing functionality can be adopted, so that the level of the test can be matched to the student's ability (Govender, 2003). Students can use automated assessment tools for practice, self-assessment and revision so as to improve their personal expertise and competencies, with a view to ultimately enhancing their examination performance. However, as Jordan (2011) states, feedback is only beneficial if the student uses it to close the gap between their current level (of knowledge) and one to which they aspire. Students indicated that electronic feedback is usually helpful, except where they disagreed with the marking, and thus ignored the feedback provided. Feedback which 'praised' the student was not appreciated. The most appreciated benefits of computer-generated feedback are that it is impersonal, objective, non-judgemental, and permits students to make their mistakes privately (Jordan, 2011).

The extent of engagement varies from formative to summative assessment, as well as from student to student. In formative assessment, some students deliberately do not answer questions, or they ignore the feedback provided, in an effort to reach the final memorandum of answers as soon as possible, with little interest in understanding the question, or the feedback offered. In general, detailed and tailored feedback is more useful than feedback which simply indicates what is right or wrong. Feedback can be made more effective if it is easily understood by students, and if it can be customised to their errors, thus enhancing their confidence in the ability of the marker, whether human or computer (Jordan, 2011).

As Duarte, Nunes, Neto and Chambel (2006: 244) state, 'feedback is the most powerful, single moderator that enhances achievement'. Thus e-assessment tools have the capability and role of providing a student with rich or detailed feedback about a specific error within an answer – possibly with counter examples, or by directing the student to a specific set of learning material (Duarte et al., 2006). Feedback should ideally enhance students' self-esteem, inspire and motivate them (Mostert et al, 2012). Thus e-assessment is seen as an appropriate tool for formative assessment, as its unique features can promote deep learning and understanding of concepts. This occurs particularly when it is used for practice.

Formative assessment has been discussed in Section 2.2.1. Formative assessment is also characterised by the distinct types of feedback it can provide to help students to improve their knowledge and learning. Useful formative feedback that could be given to students includes:

- feedback immediately after each question rather than at the end of the test,
- the score obtained on each test item,
- the final score of the test together with a breakdown of the marks obtained for each content area,
- a comparison of the correct answer(s) with the student's answer(s), and
- the model answer (Mostert et al, 2012).

Most importantly, timely (if not immediate), and constructive feedback help to motivate students effectively.

A study by Whitelock, Gilbert and Gale (2011) indicated that in traditional forms of assessment, some of the institution's examination results to students were delayed by up to a period of two months. However, with the introduction of e-assessment, students were provided with almost immediate feedback.

Traditional feedback is rarely consistent; however, e-assessment systems can support students by sending them individualised feedback through comments that are customised to their individual assessment (Tsintsifas, 2002).

2.5.7.7 Time/space independence

The inherent nature of e-assessments makes them available on demand, often independent of time and place, 'any time any place' (Cook & Jenkins, 2010: 3) both for students to take assessments and for educators to review or create them. Students have the opportunity to complete certain assessments at their own time and convenience, thus breaking the time constraints of in-class assessments (Douglas & McGarty, 2001). Continuous availability is possible through the administration of e-assessments via computers in offline settings, in network configurations, or over the Internet (Byrnes & Ellis, 2006). In such cases, students can attempt certain assessments whenever they choose, and as often as they wish (Catley, 2004). It must be noted that this research focuses on e-assessments in a controlled environment, as explained under limitations in Section 1.7.2, hence the security issues of uncontrolled environments are not addressed.

2.5.7.8 More frequent testing without increased marking overhead

e-Assessment offers an opportunity for academics to frequently assess their students without increasing their marking commitments (Maurice & Day 2004), as outlined in Sections 2.5.7.4 and 2.5.7.5. It provides the educator with an opportunity to gauge students' understanding of material in a quick and efficient manner (Souali et al., 2011). Thus, e-assessment is ideal for formative assessment, since the process provides both the educators and the students the facility to recognise problems and enhance learning (Byrnes & Ellis, 2006). Educators are able to determine the difficulties faced by the students and put more emphasis on the corresponding sections of the course material. This can result in learning gains, especially by helping low-scoring students to improve their marks in future assessments; reducing the range of scores overall in a learning module; and raising the overall performance of students within a module (Byrnes & Ellis, 2006). e-Assessment can also promote adaptive testing by presenting the student with an individualised assessment that has been automatically adapted by the system, based on the student's performance on that particular material in previous assessments (Cook & Jenkins, 2010). This allows students to retake assessments which focus on the learning material in which they are currently weak.

2.5.7.9 Question banks and formats

A wide range of questions on a single topic can be created over time, for inclusion in a question bank (Cook & Jenkins, 2010). These questions can be randomly generated per assessment, as presented in Section 2.5.7.2, using a generalised algorithm that provides the flexibility to offer a wide variety in the kinds of questions in a single assessment. Once the question banks are created, administration of the tests is simplified. Repeated use of questions reduces the time required to develop and analyse tests (Karl et al., 2011). Further, e-assessment systems can be designed to be modular in nature, so that the assessments and items can be reused and recombined to make varying assessments (Duarte et al., 2006; Cook & Jenkins 2010). Customised questions can measure specific module, learning and course objectives (Douglas & McGarty, 2001).

e-Assessment offers the potential to introduce animated graphics and multimedia into questions, which is not possible with paper assessments. With the integration of multimedia features, including video and audio, in most e-assessment software, a wider variety of question types can be created. The use of features and formats that are not feasible in the manual testing approach, range from simple adaptations of multiple choice items to highly innovative item types (Byrnes & Ellis, 2006; Dempster, 2007). The most attractive feature of an e-assessment tool is the ability to create questions in a wide variety of formats – such as hotspot, fill in the blanks, drag-and-drop, free-text entry of numbers, letters and words, free-text phrase or sentence of up to 20 words, true/false, ordering, matching, free-answering mathematical exercises, and supplementary questions. (Dempster, 2007; Duarte et al., 2006; Jordan, 2011). This is addressed further in Section 3.1.5.

The ability to deliver questions in random order, or to include ‘jumps’, as well as providing a facility to retry an attempt at a question, enhances the delivery process of e-assessments (Dempster, 2007). Reusability is another key feature supported by the question banks in most e-assessment software. This allows educators to reuse exercises built into the tool. Reused questions can be used for either formative or summative purposes (Duarte et al., 2006).

2.5.7.10 Cost savings

For universities/academics battling with restricted budgets, the use of e-assessment can greatly reduce printing costs, despite the increase in student numbers. Besides the savings on paper achieved, additional cost savings are made on the number of personnel required to mark assessments and administer their delivery (Byrnes & Ellis, 2006).

2.5.7.11 Holistic evaluation of courses

The recording and management facilities of e-assessment tools can also be used to provide educators with module evaluations (Byrnes & Ellis, 2006). This can be undertaken by analysing performance in a comprehensive set of assessments over the duration of the course. Analytics extracted from e-assessments can be used to support learning, since students can use these analytics to study their own progress, and also compare it against other students, for example, by noting class averages (Jordan, 2013).

2.5.8 Disadvantages of e-assessment

It must be acknowledged that, despite widespread adoption of e-assessment, there are also disadvantages associated with its implementation, which include issues related to security, venues, use and usability, HOTS (higher order thinking skills), culture and organization, negative or partial marking, and system complexity, as presented below:

2.5.8.1 Security

Data security can be problematic. The test taker's identity cannot always be accurately determined, which makes it possible for a substitute to take a student's place (Engelbrecht & Harding, 2003; Khedo, 2005). Moreover, management problems and logistical errors can occur, such as students forgetting their passwords, and the Internet not being available (Miller, 2012).

2.5.8.2 Venues

Computing facilities where students can access the technology to take the assessment are required. This can prove problematic for students undertaking distance-learning. However, in the context of this research, contact-learning institutions usually have computer laboratories which can be utilised (Conole & Warburton, 2005; Fielding & Bingham, 2003; Hepplestone & Helm, 2003). Accessibility to these venues for the physically challenged must also be considered (Maurice & Day, 2004; Singh & de Villiers, 2010).

2.5.8.3 Use and usability

Programs that present tests and examinations to students should be user-friendly and have a high level of usability (Walker, Topping & Rodrigues, 2008). If students struggle with the mechanics of the software, they will be distracted from concentrating on their responses. In addition, if interfaces are complicated and unfriendly, students with poor IT skills or who dislike the delivery method, may be disadvantaged (Hepplestone & Helm, 2003; Singh & de Villiers, 2010). Some e-assessment tools have limited features for the disabled (Maurice & Day, 2004; Singh & de Villiers, 2010).

2.5.8.4 Difficulty in assessing higher order thinking skills (HOTS)

The creation of MCQ questions that assess higher-level thinking can prove to be time-consuming and difficult for the educator (Miller, 2012). Most often, adoption of e-assessment is simply to measure students' knowledge, skills, and aptitudes and to rank students – thus mainly adopting objective questions (Cook & Jenkins, 2010). However, e-assessment can encourage guessing, as a student may answer a question correctly not because he or she knows the answer, but simply because he or she has guessed the correct answer from the options provided. Research has shown that in an assessment consisting of 100 multiple choice questions with five options per question, a student who has not attended any lectures and has not studied any of the material might still get several answers correct simply by guessing (Souali et al., 2011). Applying negative marking in an assessment can help to discourage guessing (Cook & Jenkins, 2010).

Therefore, e-assessment raises concerns about its capacity to assess higher-order learning (Kuechler & Simkin, 2003; Lowry, 2005; McCoubrie, 2004). Developing good MCQs is a skill and it takes time to develop valid test items. Students tend to use low-level cognitive skills, such as memorising facts and identifying correct answers from the options provided, rather than showing critical thinking and reasoning in their responses. They do this because they are expected to 'converge upon the right answer and not to diverge on a range of possibilities which a question may open up' (Souali et al., 2011: 3). Students may then become comfortable in narrowly reproducing the material taught, rather than developing higher-order cognition abilities of synthesis and evaluation. Although e-assessment is a valuable tool for both formative and summative assessment, especially if the educator develops a large question bank which facilitates the random selection of questions per assessment, the questions created must be of a high quality so that they can be reused. It is often a time-consuming task to develop good quality questions (Cook & Jenkins, 2010; Mostert et al, 2012). The issue of using MCQs to assess higher order thinking skills (HOTS) is addressed in Section 3.1.3

2.5.8.5 Culture and organisation

At some institutions, cultural and organisational barriers may exist, especially in cases where e-assessment does not fit naturally into the existing organisational structures. At times there are also political implications which may conflict with institutional structures. Furthermore, policies that force academics to maintain assessment traditions, can also act as a barrier to the effective implementation of e-assessment (Khedo, 2005; Souali et al., 2011).

2.5.8.6 Negative/partial marking

Responses to questions of the MCQ genre are judged either correct or incorrect. In some cases, the consequence of an incorrect answer is so-called negative marking, that is, the deduction of marks to prevent guessing. A further problem is the inability of e-assessment to give partial credit, which is a feature lacking in most e-assessment tools (Lambert, 2004).

2.5.8.7 System complexity

e-Assessment systems sometimes require complex technical skills and time to install them successfully (Miller, 2012). They may require a large amount of computer infrastructure which has to be installed, following specific procedures (Cook & Jenkins, 2010). Educators may find it easier to use a system that is familiar to them, even if it is a manual one. On the other hand, e-assessment systems that are relatively inexpensive and that provide an easy-to-use interface and technical tools to develop the tests, are more readily adopted (Miller, 2012). Training courses and post-installation support also attract non-technical educators.

2.5.8.8 Cost

The high cost associated with the implementation of e-assessment is another key disadvantage associated with its adoption (JISC, 2007). To be fully integrated into any institution, e-assessment tools must provide mutually compatible interfaces which adhere to universal technical standards. This increases the costs associated with e-assessment implementation, thus impeding its growth.

2.5.8.9 Power reliance

Computer systems cannot be used during power outages. The implementation of power backups is expensive and, in most cases, is subject to lead time and planning (Mogey, 2011).

2.5.9 Constraints associated with e-assessment

In implementing e-assessment, the educator and/or the administrator must consider certain constraints, as outlined below. These include training and ease of use, feedback format and timing, question readability, and the connection to course material.

2.5.9.1 Training and ease of use for students

For e-assessment to be successful, as with traditional assessment, the material on which students are to be assessed, must be available to them timeously. The assessment itself must be focused on current learning material so that students can adequately prepare for the assessment. Further, the software tool must be easy for the students to use so that they spend more time answering the assessment rather than working out how to use the tool (Alessi & Trollip, 2001; Khan, 1997).

2.5.9.2 Feedback format and timing

Ideally, feedback on an assessment should be provided at the end. Piecemeal feedback increases the duration of an assessment, and this may annoy well-informed students. It is also stated that immediate feedback directly after a question, can disturb students' focus when they are required to answer a set of closely related questions. Further, if the test has a time limit and piecemeal feedback is provided, time must be allocated for the student to read and understand the feedback (Alessi & Trollip, 2001; Horton, 2000).

2.5.9.3 Question readability

Questions must be presented simply so that students can focus on their actual response to each item, rather than spending unnecessary time deciphering how they are expected to answer the questions presented to them (Davies & Gupta, 2001; Horton, 2000; Singh & de Villiers, 2010).

2.5.9.4 Connection to course material

Course outlines are imperative as they present clearly defined outcomes, which students can use to draw a relationship between their learning and the assessments linked to this learning. The purpose and content of the assessment must be clearly explained to the students. The assessment is more valuable to the students if each item in the assessment is clearly related to a stated module objective (Singh & de Villiers, 2010).

2.5.10 Overcoming problems/drawbacks associated with e-assessment implementation – solutions to the constraints

Various guidelines on reducing the problems associated with the adoption of e-assessment are outlined below. These include training and support, practice sessions, clarity and transparency, planning, pilot testing, secure testing facility, security management, and repeated attempts.

2.5.10.1 Training and support for educators

Training should be given to educators on the design of questions for e-assessment. Offering a training session for interested staff helps to create a significant change in the mindset and attitude of those wanting to adopt e-assessment. Many academics are used to traditional, paper-based assessment and therefore hesitant to adopt any form of e-assessment (Fielding & Bingham, 2003; Horton, 2000; Khedo, 2005; Messing, 2004). If training is offered, besides providing guidelines on adopting the tool, users should also be empowered with the skills required to maintain a reliable and robust e-assessment system. This training program should be a compulsory precondition for any staff wishing to adopt e-assessment, especially for summative assessment (Khedo, 2005; McAlpine, 2002). A group with interested representatives from all the departments likely to adopt e-assessment can also help to create a positive impact on users' adoption of these tools (Khedo, 2005; McAlpine, 2002).

2.5.10.2 Practice sessions for students

Giving students sufficient time to acquaint themselves with the system before they are required to use it for summative assessment, will help to reduce students' hesitancy. In addition, providing supporting documentation to students, will also help familiarise them with the e-assessment system, without human assistance. This also helps to ensure that those with low level IT skills are not disadvantaged (Fielding & Bingham, 2003; McAlpine, 2002). With the use of practice tests, students are able to familiarise themselves with the system to be adopted for assessment (Miller, 2012). This is further elaborated in Section 2.5.10.8.

2.5.10.3 Clarity and transparency

Students need to be made aware of 'how they will be assessed, what subject areas and learning outcomes are involved, and what criteria will be used' (McAlpine 2002: 28). Thus, it is important to ensure that the intended learning outcomes of the course are aligned with the role and purpose of the assessment designed. Moreover, allowing students to monitor a list with all their test scores, gives them confidence both in the tool and in the marking system (Miller, 2012). During the class, pointing students to material that could possibly appear in their assessments, helps to motivate their learning and preparation for the assessment (Miller, 2012).

Printing facilities for students to print their assessment results, allows them to share their learning with others – such as their parents – providing transparency in the learning process (Miller, 2012).

2.5.10.4 Planning

Planning in advance for e-assessment requires close collaboration between educators, IT support and administrative staff. This will ensure that the quality and quantity of the IT infrastructure and support is adequate for the assessment (Khedo, 2005; McAlpine, 2002).

2.5.10.5 Pilot testing

The introduction of an e-assessment system on a pilot basis (even a series of pilots) prior to rolling it out on an operational basis will prove invaluable for important summative assessments. Piloting provides a vital opportunity both to test operational procedures and gain feedback from stakeholders – educators and students, prior to full implementation of the system (Fielding & Bingham, 2003; Khedo, 2005; McAlpine, 2002).

2.5.10.6 Secure testing facility

A central venue, where formal and summative e-assessment can be delivered, should be established in cases of contact teaching, where all the testing occurs at a stated site/s. This will ensure that controlled conditions can be created. Security should be ensured by predetermining who has access to the assessment while it is 'live'. This will ensure the security of the files used (Fielding & Bingham, 2003;

Messing, 2004). Assessments should require the presence or availability of designated academic and technical staff to respond to unforeseen academic or technical problems that may occur during the live assessment (Messing, 2004). Maintenance checks of all hardware and software are essential. Should any repairs or software updates be required, they must be completed prior to the actual assessment taking place. Contingency arrangements should be available for invigilators in the event of workstation/server failure. These should include guidelines on when to try and restart the system, when to transfer students to other computers, whether to end the assessment, or when to provide paper copies of the test or examination (Khedo, 2005). Additional computers should be made available in the event of failure. Similarly, extra supplementary equipment such as mice, keyboards and cables, should be readily accessible (Khedo, 2005; McAlpine, 2002).

2.5.10.7 Security management

The time period during which the assessment is 'visible', must be limited. There should be a capacity to audit individuals to verify those who logged on against those who actually participated in the assessment (Fielding & Bingham, 2003). Following the assessment, safe storage for the results and the questions, should be created, just as is done with traditional paper-based assessments. Responsibilities of academic, administrative, and support staff should be clearly defined in this area (Khedo, 2005; McAlpine, 2002; Messing, 2004). Allocating default passwords to students, and allowing them to change their passwords only when they are sufficiently confident with the system, increases security of the tool. Practising logging on and off with the password also assists (Miller, 2012). Setting a time when the assessment opens and defining which groups or individuals can access it, facilitates the security of an assessment, especially if the same assessment has to be made available to others (such as absentees) at a later date and time (Miller, 2012).

2.5.10.8 Repeated attempts

The concept of unlimited attempts, until the student achieves 100%, is a useful way of introducing the students to material required as a prerequisite to a module. Since there is no pressure to obtain the 'pass score' immediately, there is a dual benefit: students are motivated to use the system, and they better absorb the material in the readings (Miller, 2012).

To overcome the constraints associated with the successful implementation of e-assessment Cook and Jenkins (2010) suggest that adopters follow a systematic cyclic process of planning and executing e-assessments as represented in Figure 2.2.

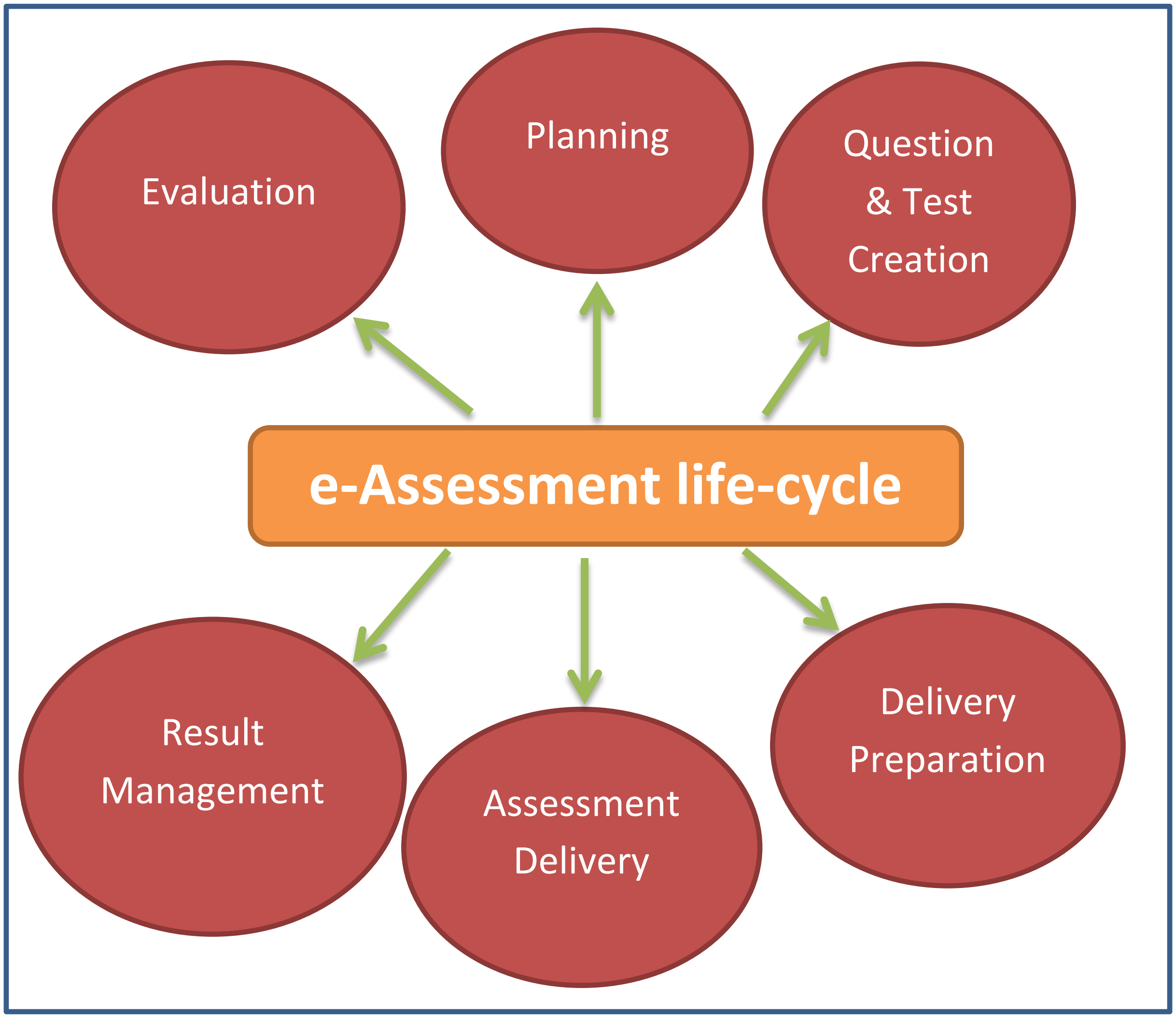


Figure 2.2: e-Assessment life-cycle
(Adapted from Cook & Jenkins (2010: 14))

2.6 Chapter conclusion

This chapter commenced with an introduction to the general literature on assessment, including discussions on its definition (Section 2.1), types (Section 2.2), purpose (Section 2.3), and measures (Section 2.4). Section 2.5 specifically focused on e-assessment, presenting comprehensive definitions (Section 2.5.1 and Section 2.5.2), the features and components of e-assessment tools (Section 2.5.3 and Section 2.5.4 respectively); common e-assessment tools adopted (Section 2.5.5) and the associated procedures when adopting e-assessment systems (Section 2.5.6); as well as the benefits (Section 2.5.7), disadvantages (Section 2.5.8), constraints (Section 2.5.9) and solutions to the constraints (Section 2.5.10) associated with its adoption.

Chapter 2 thus lays a foundation for this research, as it first considers assessment in general, then introduces e-assessment to set the context. While Chapter 2 focused on e-assessment in general, Chapter 3 homes in on the particular genre of e-assessment addressed in the research, namely multiple choice questions. In particular, it discusses the various types of MCQs supported by e-assessment tools and gives examples. Chapter 3 culminates in a synthesis of criteria from the literature that can be used in the evaluation of e-assessment tools/systems.

CHAPTER 3 Literature study: multiple choice questions (MCQs) in e-assessment

Chapter 2 was a literature study that commenced with a discussion on assessment, before specifically focusing on e-assessment. Related definitions, features, components, benefits, and disadvantages and constraints associated with e-assessment were presented. Thereafter, a brief discussion was presented on the international adoption patterns associated with e-assessment. One of the key features associated with e-assessment, as outlined in Section 2.5.3, is that e-assessment primarily adopts multiple choice questions (MCQs) for the assessments designed.

Hence, this chapter concentrates on this feature – MCQs. Section 3.1 is dedicated to detailed aspects of these questions. In Sections 3.1.1 and 3.1.2 respectively, the benefits and drawbacks of MCQs are overviewed. Since MCQs are sometimes regarded as suitable only for lower level thinking, Section 3.1.3 discusses how MCQs can be created to test higher order thinking skills (HOTS). A comprehensive discussion on the varying types of MCQs, together with examples of each, is presented in Sections 3.1.4 and 3.1.5.

Section 3.2 shifts the focus to the criteria essential for inclusion in a framework, for evaluating e-assessment systems, under consideration for adoption or evaluation. This framework subsequently provides the foundation for the fourth study in the series of studies conducted in this research (See Chapter 6, Section 6.1) which contribute to the development of SEAT and finally e-SEAT.

3.1 Multiple choice questions (MCQs)

e-Assessments are dominated by standard multiple choice questions (MCQs), which generally have the format of a prompt followed by a small set of responses from which students are expected to select the best option (Scalise & Gifford, 2006). The prompt/test item consists of a question or stem, the correct answer, and a set of distractors. MCQs have proven to be efficient in measuring students' achievement and are adopted internationally, both for assessment and diagnostics (Mitkov, Ha & Karamanis, 2006). The reason for the widespread adoption of MCQs is threefold: they should efficiently cover the educational content; they have a high degree of reliability; and they are easy to score (Kadhi, 2004;

Roberts, 2006). The previous chapter pointed out that e-assessment can play valuable roles in both formative and summative assessment, and this is particularly true for MCQs.

MCQs offer the best assessment tool for large numbers of students and where limited resources are available to educators (Bani-Ahmad & Audeh, 2010), since a large number of tests can be corrected automatically (Costagliola & Fuccella, 2009) within a short period of time (Mendes, Curto & Coheur, 2011; Ventouras et al., 2010). Early adoption of MCQs took place in medical assessments and frequent use occurs in computer programming courses (Byrnes & Ellis, 2006). Medical education was a pioneer of electronic testing, due to the fact that as students progress in their medical studies they spend more time on practical procedures off-campus, and it becomes increasingly difficult to gather them in a single venue for a test. More recently, MCQs are widely adopted in medical, dental and allied health education due to their ability to assess a broad range of knowledge in a short period of time, and are thus favoured over methods such as short answer, essay and oral examination formats (Ware, Kattan, Siddiqui, & Mohammed, 2014).

Multiple choice testing is recommended for testing factual recognition, which represents only the basic level of professional competence. As a consequence, standard multiple choice testing has limited applicability in modern competency-based education (Karl et al., 2011) and in assessing critical thinking. Nevertheless, standard MCQs are commonly adopted in many domains and are used as primary assessment tools (Beullens, van Damme, Jaspaert & Janssen, 2002). However, newer and improved MCQ formats and techniques have emerged (Osika, 2006; Prestera, Clariana & Peck, 2005) and are addressed in detail in Section 3.1.5, along with examples.

If constructed properly and written well, MCQs can be an effective tool for assessing skills and knowledge in students. Written badly, they can be confusing and demotivating (Alton, 2009). Educators generally find that it is more difficult to create MCQs, despite the range of styles of MCQs that can be adopted. This difficulty is often due to inexperience. Furthermore, the process of generating plausible, yet definitely wrong, distractors is complex and time-consuming (Mitkov & Ha., 2003).

MCQs offer various advantages in addition to the benefits of e-assessment outlined in Section 2.5.7. Section 2.5.7 considered the benefits of e-assessment in general, while the next section, Section 3.1.1, is dedicated to MCQs in particular. Similarly, Section 3.1.2 addresses the distinct drawbacks of MCQs.

3.1.1 Benefits of MCQs

MCQs offer the possibility of covering a broad set of topics (Mendes, Curto & Coheur, 2011), since the questions are usually short.

Although writing good quality MCQs for e-assessment can be time-consuming, benefits received are not just in terms of student performance and automated marking in routine assessment, but also in time saved by academics released from creating and arranging aegrotats or supplementary assessments for students who did not take a test or examination due to a valid reason.

MCQs can be drawn from standardised item banks, which in some disciplines and topics, are freely available, but these questions may need to be contextualised to meet the needs of the students' local linguistic features and the concepts being assessed (Luckett & Sutherland, 2000). In most cases, the question banks are custom-built by the educators presenting a particular module.

Item analysis and item response theory (IRT) allow educators to evaluate the quality of their MCQs in terms of difficulty and discriminative capacity (Costagliola & Fuccella, 2009). Provided that MCQs have been correctly formulated by the educator, selection of the correct option requires specialised knowledge on the part of students, a grasp of detail, quick responses, and decision-making skills, taking into account that, in some cases, specified time durations might be predetermined for answering the set of questions (Ventouras et al., 2010). Well-designed MCQs offer the further cognitive advantages of lending themselves to the verification of knowledge, comprehension of concepts, and achievement of course objectives (Costagliola & Fuccella, 2009).

After the assessment, the student may print the results, see his/her final score and his/her mistakes, if adequate feedback is provided (Ventouras et al., 2010). Students appreciate timely feedback that is well-presented and simple to understand. This feedback assists their revision for further study (Malau-Aduli, et al., 2013).

Scoring/marking of assessments of MCQ format is objective (Ventouras et al., 2010), since it is free from bias and distortional effects, such as emotional judgments (Costagliola & Fuccella, 2009). Ventouras et al. (2010) indicate that multiple choice questions (MCQs) provide higher reliability and are as valid as constructed-response questions.

Essentially, they are useful for quickly identifying a student's understanding of a field, but are also useful for revision purposes via formative assessment, because of the speed and accuracy with which they can be assessed, especially through automated marking systems (Elstein, 1993; Farthing, Jones & McPhee, 1998).

3.1.2 Drawbacks associated with MCQs

Despite their widespread adoption, the value of MCQs in educational spheres has often been criticised as they are viewed as tests of factual recall, as has been mentioned in Section 3.1.2. Many MCQs do not assess application of knowledge for problem solving, partly because of the way in which questions are constructed. Due to the somewhat artificial way in which they are marked, whereby a single answer needs to be selected, this form of testing is considered to be unnatural (Engelbrecht & Harding, 2003). In the real world, one is never faced with a problem and five possible solutions, with one viewed as exclusively correct. Furthermore, MCQs have been widely criticised as being artificial and unrelated to authentic practice, since reasoning and problem-solving skills are difficult to measure in MCQ format. According to Fenderson et al. (1997: 526), they tend to focus on 'recall of trivia'.

Although MCQs have been adopted since the early 1900s, considerable skill, care and practice is required when writing them, to avoid confusing students with unanswerable questions or poor alternative answers (Alton, 2009). Engelbrecht and Harding (2003) point out that if a question is not clear to students, they may select an incorrect option due to a misleading question and not due to insufficient knowledge. One of the main challenges in constructing the MCQ test item is the selection of plausible distractors which will better distinguish confident test takers from unconfident ones (Mitkov et al, 2006).

Incorrect selection is not always the result of a student's lack of knowledge or understanding, or confusion. It is possible that students have a correct understanding but make minor errors in selection

(Fenderson et al., 1997). Hence, some researchers hold that MCQs should make provision for partial credit, without treating minor and major mistakes as equal. A word of caution though - 'if an assessment always provides credit for partially correct answers, students can pass the entire module without having understood any concepts fully' (Engelbrecht & Harding, 2003: 57).

Despite the convenience offered by the MCQ format, such assessments can be regarded as low on validity if they assess trivial knowledge only (Luckett & Sutherland, 2000).

MCQ assessments sometimes encourage 'poor attitudes toward learning and incorrect inferences about its purposes … for example, that there is only one right answer, that the right answer resides in the head of the educator or test maker, and that the job of the student is to get the answer by guessing' (Bleske-Rechek, Zeug & Webb, 2007: 94). Some theorists argue that MCQs presume that complex skills can be decomposed and decontextualised, which is not always possible. Rather, MCQs rely on well-structured problems with algorithmic solutions. Thus students have the impression that knowledge is additive rather than integrative (Bleske-Rechek, Zeug & Webb, 2007; Scalise & Gifford, 2006).

MCQs pose a limitation on the kind of questions that might be adopted. The nature of MCQs judges the student solely on the correctness of the answer he/she chooses and not based on the method used for reaching the answer (Ventouras et al., 2010). Moreover, they do not allow the educator to investigate in depth, whether the topic which a specific question addressed, has been fully understood or not (Ventouras et al., 2010).

Guessing allows a student to obtain partial scores in the final score, by answering questions by chance, without possessing knowledge of the questioned material. By guessing, it is usually possible to get some questions right. Without negative marking, students gain marks for correct answers and lose none for omissions (Ventouras et al., 2010). Certain systems offer an option for mixed-scoring - that is, negative and positive marking – whereby students gain marks for correct answers and also lose marks for incorrect answers. A study by Ventouras et al. (2010) showed that students are less willing to answer such questions when compared to MCQs based only on positive scoring rules (Bleske-Rechek, Zeug & Webb, 2007). Thus mixed-scoring rules might induce a 'hampering' effect to the student, dissuading him/her from tackling a question for which he/she may possess an intermediate level of knowledge.

Writing good MCQs is not an easy task. Creation of good quality questions that assess more than simple learning outcomes can be a time-consuming and labour-intensive process for educators (Mendes, Curto & Coheur, 2011; Pittenger &Lounsbery, 2011). In addition, the updating of questions in the MCQ bank requires much time and effort and often results in inconsistent quality when questions are provided by different educators (Pitenger & Lounsbery, 2011).

Multiple choice assessment approaches are often criticised for not facilitating active learning, because they provide students with a list of choices rather than requiring them to actively identify the correct choice and explain or justify why it is best, as they would be required to do in real life or in a written assessment (Pittenger & Lounsbery, 2011).

3.1.3 MCQs for higher order thinking skills (HOTS)

Research shows that it is possible to test higher-order thinking through well-developed and researched MCQs, but this requires considerable skill, practice and time on the part of the educator (Luckett and Sutherland, 2000; Mitkov & Ha, 2003). As a result of the extra time taken to write high quality MCQs, some educators tend to administer less e-assessments than anticipated in their planning (Catley, 2004).

e-Assessment is often thought of as solely utilising quiz tools (Mostert et al, 2012), implying that it is best suited for recall-type multiple choice type questions. However, the application of MCQs to assess higher order thinking skills (HOTS) is becoming increasingly common. Although in most cases MCQs focus on factual knowledge, they may discerningly be designed to assess HOTS, as well as knowledge. Mostert de Bruyn and Pretorius (2012: 18) indicate that at the University of Pretoria (UP) in South Africa, ‘it has been proven over a period of approximately 20 years … that e-assessment can be used effectively to enhance student learning through assessing on higher level cognitive levels of Bloom’s taxonomy, in different disciplines’.

Essentially, MCQs, in whatever format, have proven to be an efficient tool for measuring students’ achievement but are best used in combination with other assessment methods (Luckett & Sutherland, 1997; Mitkov et al, 2006).

3.1.4 Types of questions supported by e-assessment tools

This section briefly discusses the most common types of questions supported by e-assessment tools. This serves as an introduction to the comprehensive discussion on the many MCQ types described and illustrated with examples in Section 3.1.5.

In general assessment, questions can be classified into two categories, namely, Constructed Response Questions (CRQs) and Provided Response Questions (PRQs).

- CRQs require students to construct their own response to questions posed. This category includes open-ended written questions, essays, projects, short-answer questions (paper-based or online), free-response and paper assignments, that is to say, the forms used in traditional assessments.
- PRQs allow students to choose between a selection of given responses, hence PRQs are best suited for e-assessment. The main criticism is that the rigidity of the marking allocations in PRQs does not allow for flexibility, for example, it is not possible to allocate a proportionate mark.

The most common question types adopted with PRQs are multiple choice questions (MCQs), multiple response questions (MRQs), matching questions, and hotspot questions. However, there is a wider variety of several question types that can be used in online tests, including: multiple choice, true/false, true/false with reason, matching, ordering, fill-in-the-blanks, completing and correcting code, and writing new code (in the context of computer programming) (Byrnes & Ellis, 2006; Costagliola & Fuccella, 2009; Souali et al., 2011). As explained in the introductory section of 3.1, these newer and improved formats and techniques will be addressed in detail in Section 3.1.5, along with examples to illustrate each type.

Most of the questions included in online tests are characterised by closed stimulus and response. Tests including only these questions are called objective tests (Costagliola & Fuccella, 2009). Stressed by many authors, e-assessment supports a variety of objective question types that can be adopted (Alessi & Trollip, 2001; Byrnes & Ellis, 2006; Costagliola & Fuccella, 2009; Engelbrecht & Harding, 2003; Fielding & Bingham, 2003; Horton & Horton, 2003; Khedo, 2005; Lambert, 2004; Maurice & Day, 2004; Sim et al., 2004; Souali et al., 2011). The various formats of MCQs are discussed and illustrated in Section 3.1.5.

The most common question type adopted in e-assessment is straight multiple choice – already described in this chapter – where the student has to choose the correct option from a list of possible answers (Souali et al., 2011). Students respond by pointing and clicking the mouse on the selected answer, by moving objects around the screen, by entering numbers or characters via the keyboard, or by pointing gestures on tablets. The software can react with an appropriate result and, frequently, textual feedback. Some sophisticated programs select subsequent questions based on the previous answers that the student gave (adaptive testing) – setting more demanding questions if answers are correct, or easier questions about the same topic if answers are incorrect (Souali et al., 2011). The incorporation of such questions can strengthen e-assessment, and is addressed in Section 3.1.3, which reports that MCQs can also be adopted for assessing HOTS. Since e-assessment implies the use of computers to deliver assessments, as well as mark and analyse students' reponses, it should therefore support MCQs as well as short-answer response questions (Mostert et al, 2012).

The key to successful implementation of an e-assessment system as part of any institution's assessment strategy, is threefold. Firstly, the system's distinct features should be utilised to their fullest capacity. Secondly, good quality questions and varying question types should be created (Mostert et al, 2012). Finally, satisfactory questions should be stored in a database for reuse.

3.1.5 Different formats of MCQs supported by e-assessment tools

In the following subsections, a brief description is given of each of the varying question types supported by e-assessment tools, followed by an example.

3.1.5.1 Multiple choice questions or multiple response questions

Both multiple choice questions (MCQs) and multiple response questions (MRQs) display a list of answers from which students have to select the most appropriate answer. MCQs require students to select one option from the list of alternatives provided. e-Assessment systems allow unique ways of presenting these questions since they can be shuffled/randomised so that each student is presented with the questions in a different order, while taking the same assessment (Mostert et al, 2012). Although these are the simplest to answer, they are often assumed to assess merely low-level learning objectives, such as addressing common errors in understanding and testing memorisation of meaningful facts and concepts (Miller, 2012). See Figure 3.1 for an example.

An expert system uses a(n)______________ to select the most appropriate response.

a) inference engine
b) decision support system
c) knowledge base
d) data source

Figure 3.1: Multiple Choice (MCQ): single response example

(1st year End User Computing module)

MRQs can allow students to select multiple options when more than one is correct. They have a choice between choosing one, a combination of a few, all, or none of the alternatives (Miller, 2012). It is similar to an MCQ, but allows for more than one option to be selected. This type of question requires students to think more than they would for an MCQ before they respond. Thus, the cognitive level of MRQs is higher than MCQs (Miller, 2012). See the example in Figure 3.2.

You have a computer that runs Windows 7. You start the computer and receive the following error message:

BOOTMGR is missing. Press Ctrl+Alt+Del to restart

You then start the computer from the Windows 7 installation media. You need to ensure that the computer successfully starts Windows 7. What are two possible ways to achieve this goal?

A. Run Startup Repair
B. Run System Restore
C. Run Bootrec/RebuildBcd
D. Run Bcedit/createstore

Figure 3.2: Multiple Choice (MCQ): multiple response example
(2nd year Software module)

3.1.5.2 Extended Matching Items (EMIs) or Extended Matching Questions (EMQs)

Extended Matching Items (EMIs) or Extended Matching Questions (EMQs) are a variant of multiple choice questions. In an EMI the student selects the best answer to a question from a list of ten to twenty options, each of which may be used once, more than once, or not at all. This variant of multiple choice questions is widely adopted in the medical education field (Beullens, van Damme, Jaspaert & Janssen, 2006). EMIs aim to test the application of knowledge rather than simple recall. The aim of assessments is to get students to apply knowledge rather than simply recall isolated facts. EMI questions take the form of a small problem or short cases called vignettes. 'In three to six sentences a medical case is described giving various details such as the patient's symptoms and the results of lab tests, and the student is asked to arrive at a diagnosis' (Wood, 2003: 2). There may be several questions about the vignette and each answer will be chosen from a long list rather than from just five options. Thus, having read and understood the vignette and the information given in it, the student selects the best answer to each question from a list of up to twenty options, each of which may be used once, more than once, or not at all (Fenderson et al., 1997; Wood, 2003). EMIs were originally developed to assess diagnostic pattern-recognition skills of physicians, but are now used more widely in the medical education field (Case & Swanson, 1993). Although EMIs have been used mainly in the discipline of medicine, they lend themselves to other subject areas as well (Wood, 2003) and are currently incorporated in assessment of a wide variety of subjects.

EMIs, also known as EMQs (Extended Matching Questions) retain the advantages of MCQ tests such as objectivity and automated marking, but also offer the following unique advantages:

- The question format aids in specifying the examination content, for instance, in the medical domain, each major ailment could be used as a theme (Beullens et al., 2002; Case, Swanson & Ripkey, 1994).
- A set of questions on the same theme, facilitates the development of several content-parallel test forms (Beullens et al., 2002; Case, Swanson & Ripkey, 1994).
- EMIs require students to solve small authentic problems rather than recall isolated facts (Beullens et al., 2002; Case, Swanson & Ripkey, 1994; Fenderson, et al. 1997). Thus educators are better able to distinguish well-prepared students from the marginal students. Furthermore, they are valuable in testing core knowledge, because the provision of cues is minimised (Fenderson et al., 1997).
- The structure of EMIs facilitates item writing: the option list flows naturally from the theme and the items from the option list. The homogeneous options and parallel items reduce technical flaws made by academics in phrasing items (Beullens et al., 2002, Case, Swanson & Ripkey, 1994). They are thus easier to prepare than traditional multiple choice tests, because there is no need for plausible distractors. They emphasize real-world problem-solving skills, and are less likely to concentrate on unimportant/low-level aspects (Fenderson et al., 1997).
- The long option list allows inclusion of all relevant options (Beullens et al., 2002; Case, Swanson & Ripkey, 1994). Thus they prevent students from answering by elimination, rather than by actually knowing the answer (Fenderson et al., 1997).

An example two EMI questions in Computing, that relate to the same vignette, is provided in Figure 3.3.

Your network contains an Active Directory domain named tailspintoys.com. The domain contains a web server named Web1 that runs Windows 2008.

You create a new site named Site1.

You need to ensure that when a user enters a Fully Qualified Domain Name (FQDN) on Site1 for a server that does not exist, a custom Webpage displays.

Which feature should you configure?

A. Authentication
B. Connection Strings
C. Default Document
D. IIS Manager Permissions
E. Management Service
F. Request Filtering
G. SSL Settings
H. Worker Processes

Your network contains an Active Directory domain named tailspintoys.com. The domain contains a web server named Web1 that runs Windows 2008.

You create a new site named Site1.

You need to prevent Web1 from accepting HTTPS URLS that are longer than 512 bytes

Which feature should you configure?

A. Authentication
B. Connection Strings
C. Default Document
D. IIS Manager Permissions
E. Management Service
F. Request Filtering
G. SSL Settings
H. Worker Processes

Figure 3.3: Extended Matching Items (EMI) example

(3rd year Networks module)

3.1.5.3 Drop down items

Drop down items are used when there are a large number of options available to the student from which to select. The options are not immediately available when the question is opened. Instead, all the options required to be displayed are added to the drop down list. This question type expects the student to open the list and select only one choice as an answer from the drop down list provided. Several questions might be asked, each of which has its correct answer in the list. An example of this type is presented in Figure 3.4 showing several different questions. Note that the full list is not visible in the figure.

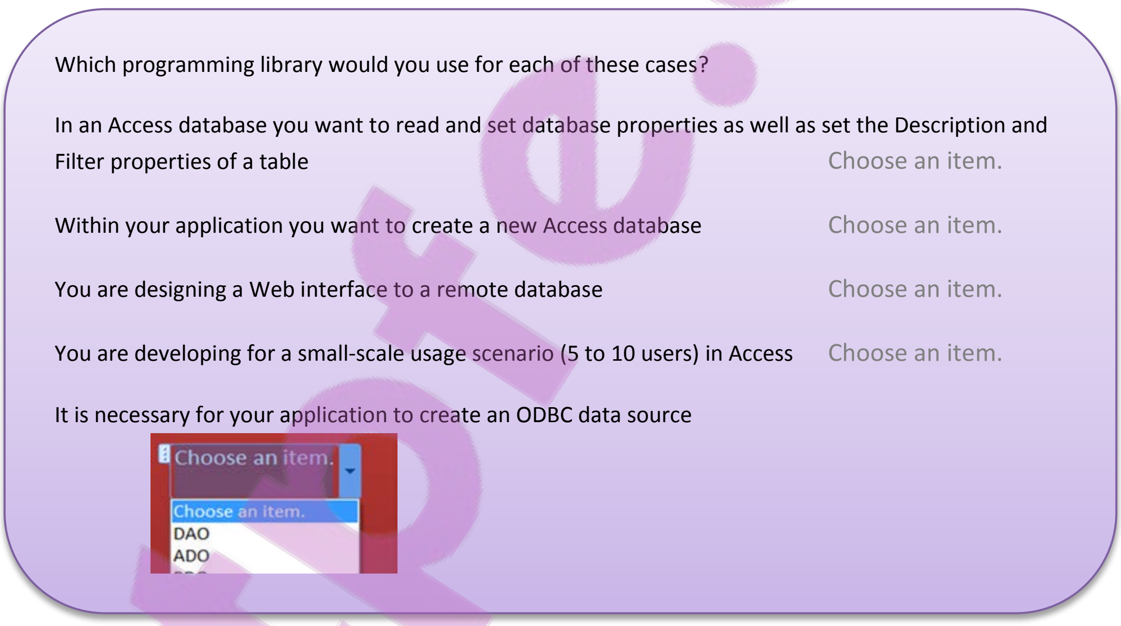


Figure 3.4: Selection/Drop Down Lists example

(3rd year Databases module)

3.1.5.4 True/False questions

True/False questions make students decide the correct answer between only two alternatives, stating whether a statement is true or false. True/False questions are easier to prepare and quicker to read and answer, so that the number of questions in a test can easily be increased and the subject matter is better sampled. Moreover, with True/False tests, it is possible to ensure that partial knowledge is restricted to 'lack of confidence' (Burton, 2001: 48). However, the possibility of guessing is an inherent concern with True/False questions. It is well-researched and findings indicate that guessing alone will give an average score of 50% with solely True/False assessments, which makes it possible for many students to pass a test, regardless of ability.

This question type has been adopted with varying degrees of success, as the educator has to ensure that every phrase (the 'whole') of the stem is either True/False (Miller, 2012). An example is given in Figure 3.5.

Management Information Systems (MIS) deal with the planning for, development, management, and use of knowledge workers to help perform all tasks related to information processing and management. TRUE or FALSE?

Figure 3.5: True/False (T/F) example
(2nd year MIS module)

An extended form is Multiple True/False questions which offer a series of statements to the student, each of which is to be judged as True/False.

3.1.5.5 True/False with explanation

The drawbacks of True/False questions mentioned above can be alleviated by requiring students to provide a reason for the selection of their answer.

As illustrated in Figure 3.6, this question type is similar to the True/False question described in Section 3.1.5.4, as it offers a statement which the student has to judge as True/False. In addition to making the judgment, however, the student has to substantiate his/her choice with a reason/explanation for the selection. The explanation is usually marked manually, while the answer selected is marked electronically by the system. These questions are reputed to be able to assess knowledge and its interpretation (Burton, 2001; Khan, Davies & Gupta, 2001).

> The average system owner is interested in raw data. [Answer True/False and give a brief reason]
>
> FALSE- Rationale: The average system owner is usually interested only in information that adds new business knowledge, rather than the raw data.

Figure 3.6: True/False (T/F) example with the correct solution and its explanation
(1st year End User Computing module)

3.1.5.6 Fill-in-the-blank questions

In a fill-in-the-blank question, students enter the word(s) that are required to complete the sentence/paragraph/table presented to them. This is assessed by comparing the word typed in by the student to the answer in the memorandum uploaded to the system. The system's correct answer typically includes alternative correct answers in the form of synonyms. See the example in Figure 3.7.

> A mathematical representation of a real-life system is a __________.

Figure 3.7: Fill-in-the-Blank/Completion example
(2nd year Systems Development module)

3.1.5.7 Hotspot questions

Hotspot questions ask the student to identify areas of an image, by using the pointing device or touchscreen to select an item or identify part of a picture. This provides the student with an opportunity to identify area(s) on a graphical image (which could include high resolution graphics and video material of live scenarios/cases). It also supports academics in assessing students on higher cognitive levels (Miller, 2012; Mostert et al, 2012). Multiple hotspot questions, where an extensive or complex photograph, diagram or model is presented, followed by a series of questions, are becoming increasingly popular. These permit the academic to get more mileage out of the hotspot base created, rather than just one mark.

A single hotspot question is illustrated in Figure 3.8.

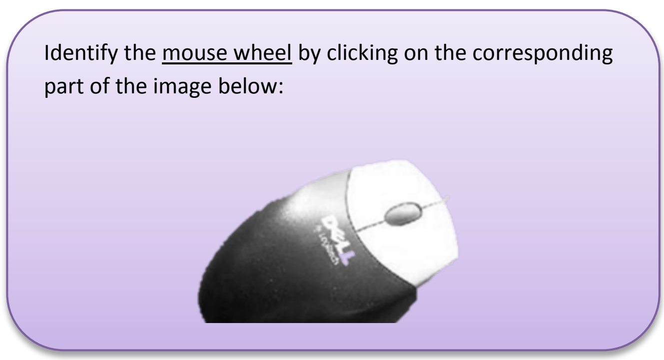


Figure 3.8: Hotspot question example

(1st year End User Computing module)

3.1.5.8 Matching list questions

Students are presented with two lists where they must identify the text or graphic items in the first list that correspond with items (text or graphics) in the second list. This structure can also be used to assess students' understanding regarding the sequence of a specific process (Mostert et al, 2012). It may be the case that not all options are used or that a particular option is the correct answer for more than one question. Hence, these questions are not only complex for the student to answer, but also challenging for the educator to set. However, they have the added advantage of assessing on a higher cognitive level (Miller, 2012). See the example provided in Figure 3.9.

Match the following:

1. Objects in a graphical user interface, ________ , are capable of storing multiple files.	**A. Font**
2. A unit of frequency, a ________, is the rate of change in the state or cycle in a sound wave, alternating current or other cyclical waveform.	**B. Encryption**
3. A ________ is the combination of typeface and other qualities such as size, pitch and spacing, that comprise a set of characters.	**C. Fiber optics**
4. Small graphical representations of an object in a graphical user interface (GUI) are called _______.	**D. Hertz**
5. A collection of similar information given a name for easy storage and retrieval is called a ______.	**E. File**
6. _________________ is the language used to write World Wide Web documents.	**F. Hypertext Markup Language (HTML)**
7. Cabling that has a core made of strands of glass or plastic is called ______________.	**G. Ground**
8. The translation of data into a code that needs to be decrypted to become legible is known as ______________.	**H. Hyperlink**
9. ______________ are the physical electronic components that make up a computer system.	**I. Expansion slots**
10. A __________ is an electrical connection with a common return for a circuit with an arbitrary zero of potential.	**J. Folders**
11. An item in an electronic document, a _________, links to another object such as a position in a document or a different document.	**K. Hardware**
12. Openings on a computer where a PCB or PC card can be inserted to add capabilities to the computer are called ______________.	**L. Icons**

Figure 3.9: Matching List example

(1st year End User Computing module)

3.1.5.9 Drag-and-drop/move-object questions

Drag-and-drop questions test the students' ability to allocate items to the required list, sometimes sequentially. This is achieved by dragging or positioning icons, images, textual labels or labels on the screen, to identify the correct areas. A strong point of these questions is that they allow the educator to test knowledge and skills that may not be possible to test on paper (Mostert et al, 2012). This is illustrated in Figure 3.10.

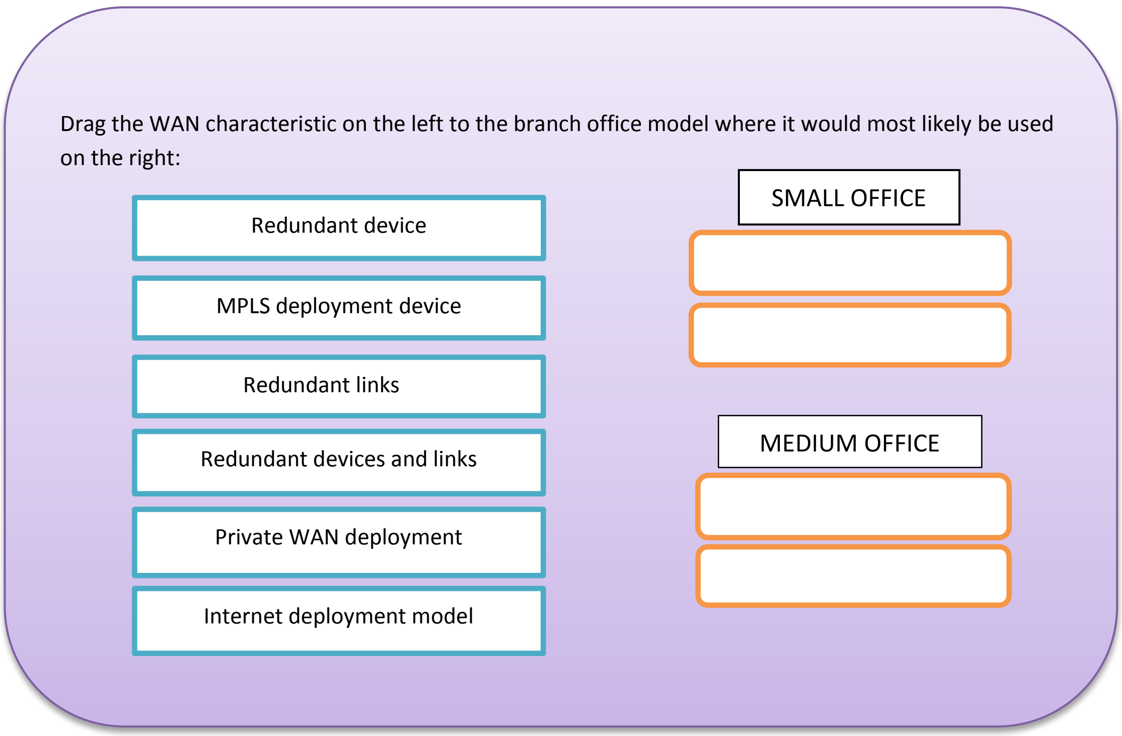


Figure 3.10: Drag-and-drop question example

(3rd year Networks module)

3.1.5.10 Diagram/Video Clips

As illustrated in Figure 3.11, students are presented with a diagram and a short explanation of the diagram. The student is expected to study the diagram and the narrative and to answer a question(s) based on both.

Video questions operate in the same way, where a student is expected to watch a short video clip and thereafter answer a question(s) based on the content related to that video clip.

A narrative is given immediately below. Consider the narrative in relation to the Entity Relationship Diagram (ERD) in **Figure A** below and answer the question below the figure.

A vet treats animals. She treats many animals per day. Sometimes, the vet is assisted by a nurse. Each animal which is treated has a treatment history – this is recorded and updated each time the animal is brought to the vet.

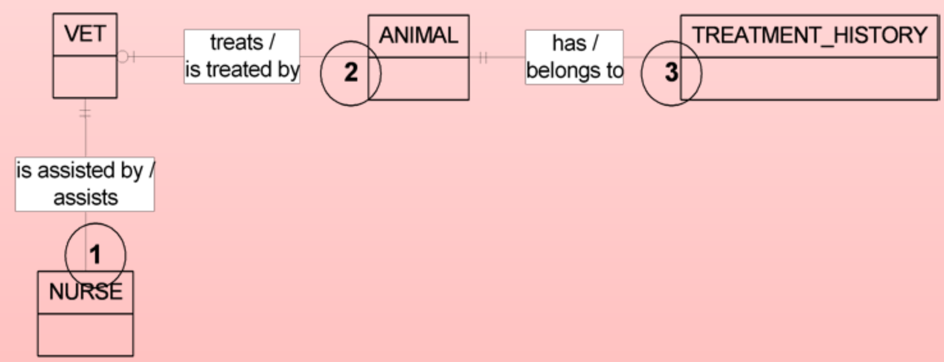


Figure A

The most appropriate notation which should replace circle 1 in **Figure A** is

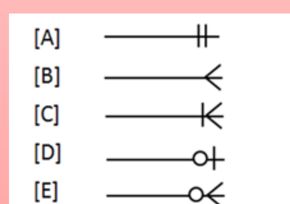


Figure 3.11: Diagram question example

(2nd Year Systems Design Module)

As is the case with the hotspot questions in Section 3.1.5.7, it is likely that several questions would be asked in relation to each diagram or video.

3.1.5.11 Simulation questions

Simulation questions are highly interactive in nature. Students are expected to perform a task on-screen, as they would in a real-world situation. This is illustrated by the example in Figure 3.12. If the correct command is entered, it will execute.

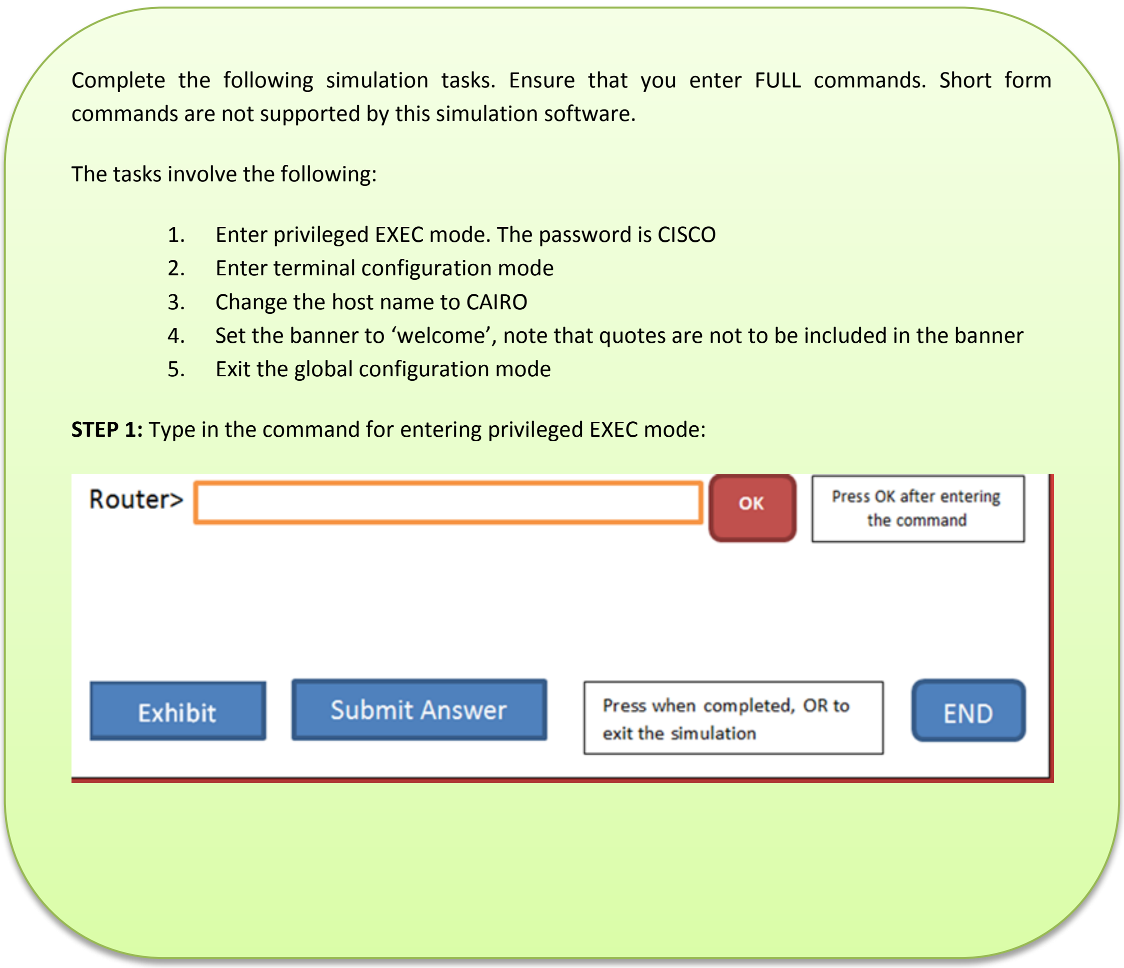


Figure 3.12: Simulation question example

(3rd year Networks module)

3.1.5.12 Ranking questions

In the ranking type of questions, students are expected to rank answers in order of importance based on a scenario or statement provided, as illustrated in Figure 3.13. Users are expected to select the options in sequence from a drop down list provided. Since sequence is important, the answers must be selected in the order required, although the same list (see Figure 3.13) is provided for all the items, A to C.

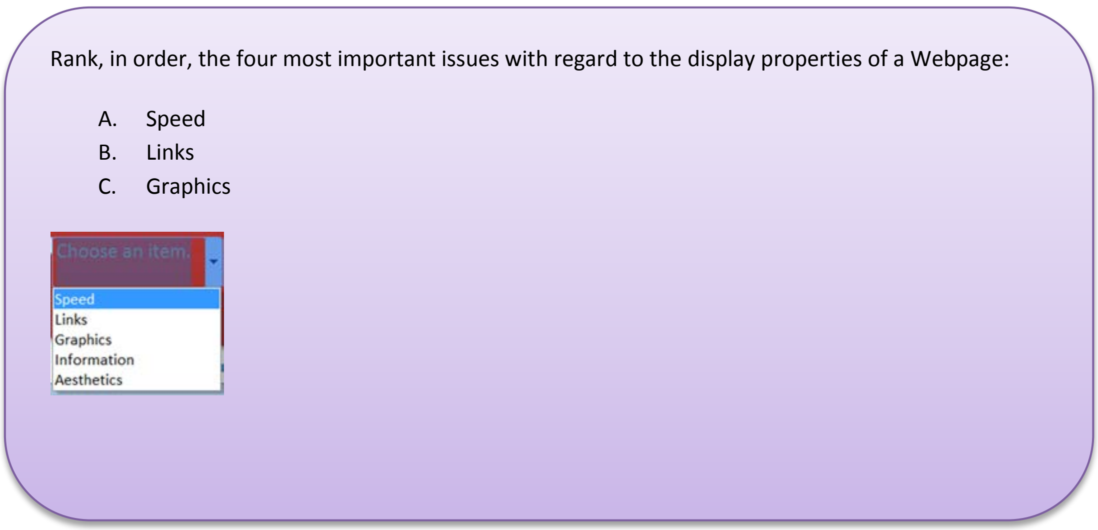


Figure 3.13: Ranking question example

(2nd year Web Design module)

3.1.5.13 Reordering/Rearrangement/Sequencing

In this type of question, the student is expected to rearrange items that have been presented in a jumbled order, into a sequence or set of items that are in order. As illustrated in Figure 3.14, the user is expected to click on the option that should appear first, and then add it to the answer area by clicking on the forward arrow. This is repeated until the required answer is completed. Should he/she need to delete an option selected, this is achieved by clicking on the back arrow.

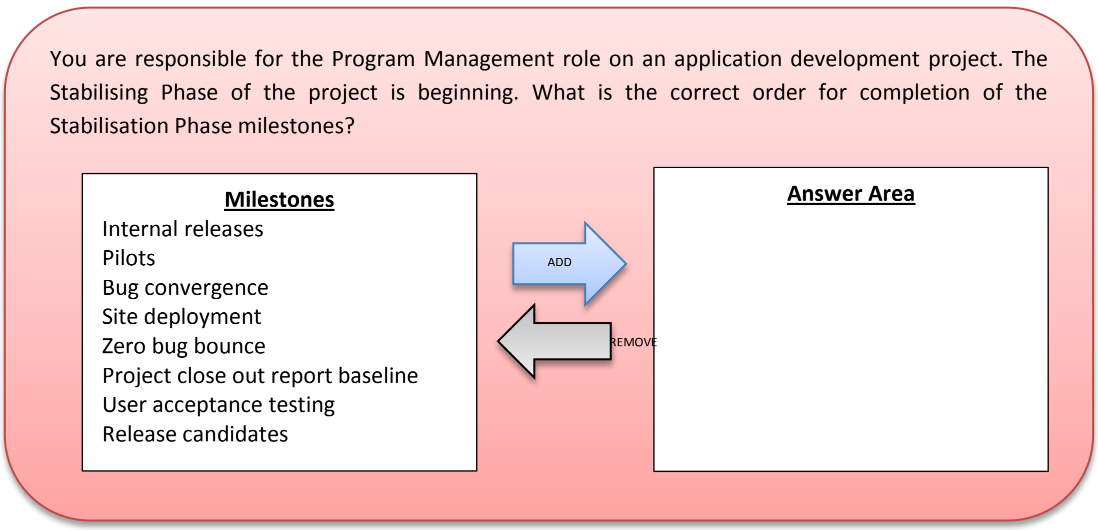


Figure 3.14: Reordering question example
(3rd year Project Management module)

3.1.5.14 Short answer questions

Short answer questions are designed such that the student has to type in the answer to a specific question. These would usually include one word or numeric character. For numeric responses educators can set ranges within which the answer may lie, or limit the number of decimal places. For words, synonyms and alternative spellings are accepted. At times the system may allow questions to be created randomly according to set parameters within specified ranges, thus providing an extended number of questions for the educator to present to the students. This is especially useful for self-assessments (Mostert et al, 2012). An example of a short-answer question is given in Figure 3.15.

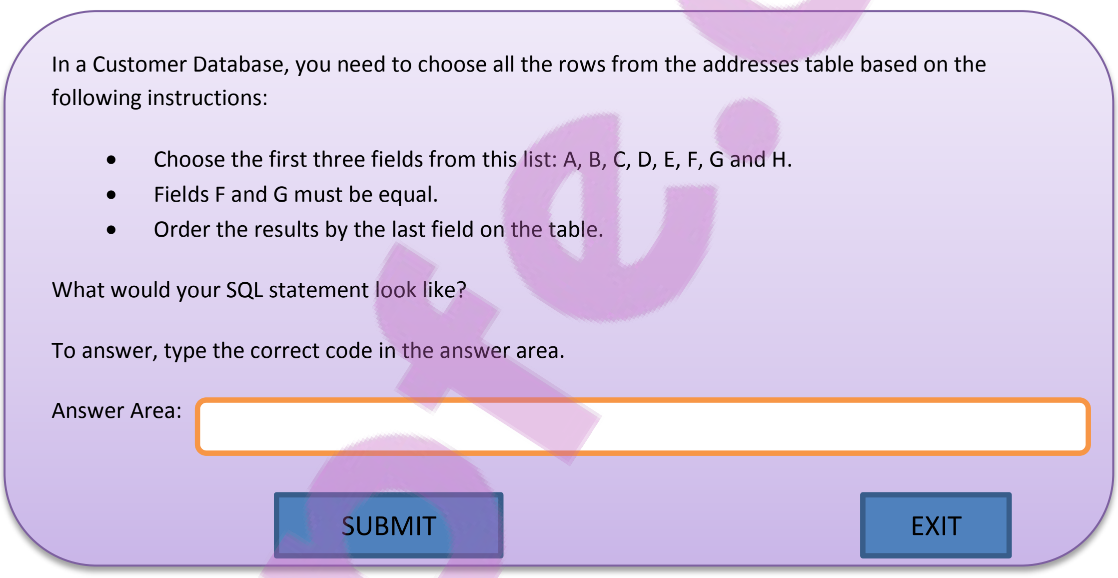


Figure 3.15: Short answer question example
(3rd year Databases Programming module)

All the question types discussed above can, in an e-assessment system, include high resolution graphics, video, sound, animations and other multimedia elements which are used as part of the questions and not just as decorative images (Miller, 2012). These questions can be adapted to test the higher cognitive levels of Bloom's taxonomy (Mostert et al, 2012).

3.2 Initial criteria for evaluating e-assessment tools/systems

A comprehensive discussion on MCQs was presented in Section 3.1. Section 3.2 discusses criteria that are appropriate for inclusion in an e-assessment system under consideration for adoption or evaluation, in an attempt to partly answer Research Question 4.

Research Question 4	***What are the requirements for selecting or personally developing an electronic assessment tool?*** ***Theory: What does the literature suggest as appropriate requirements for electronic/online testing and assessment tools?***

This discussion serves as the foundation for the development of the SEAT Framework in Study 4, in the iterative series of action research studies conducted in this research.

In South Africa, many academic institutions are faced with large student groups, hence the workload of educators is demanding. As a result of this, the use of e-assessment has increased steadily over the past ten years so that Computer-Based Testing (CBT) has become an integral part of the assessment strategy of many academic departments (Pretorius et al, 2007).

More than a decade ago, Valenti, Cucchiarelli, and Panti (2002) and Scalter and Howie (2003) stated that very little research had been carried out regarding requirements for CBT systems. This was subsequently supported by Pretorius et al. (2007), who reported that in South Africa too, inadequate information exists on the criteria required for evaluating e-assessment systems. Valenti et al. (2002) also suggest that any e-assessment system should be evaluated prior to its adoption. They divide e-assessment systems into two major components – the Test Management System (TMS) and the Test Delivery System (TDS). They explain that the TMS is designed primarily to assist the educator in creating questions and tests. Its secondary role is to provide assistance in evaluating the tests created. The TDS serves as a facilitation role, assisting in the administration and delivery of assessments to students.

Valenti et al. (2002) categorise the criteria they deem necessary for evaluating an e-assessment system into four categories, namely – *Interface, Question Management, Test Management* and *Implementation Issues*. Three of these criteria, *Question Management, Test Management* and *Implementation Issues*, are included in the list presented by Pretorius et al. (2007). Where Pretorius et al. (2007) included a *Technical* category, Valenti et al (2002) identified a fourth category, unique to their study, as the *Interface* category.

Pretorius et al. (2007) were the pioneer researchers in South Africa to identify features and attributes that are essential to any good e-assessment tool. Their criteria were grouped into four categories – *Technical, Question Management, Test Management and Implementation Criteria* – which can be adopted when evaluating or selecting a new e-assessment system. The complete range of criteria compiled by Pretorius et al. (2007) includes pre-criteria (prior to the system being used) as well as post-criteria (after using the system).

The above-mentioned categories outlined by both Valenti et al. (2002) and Pretorius et al.(2007) will be briefly discussed in Sections 3.2.1 to 3.2.5, along with individual criteria extracted from Carter et al. (2003), Fremont and Jones (1994), Lewis and Sewell (2007), and Maurice and Day (2004).

3.2.1 Category 1: Technical criteria

.

A good e-assessment system should be able to run on a variety of software *platforms* or, better still, be *platform-independent*. This facilitates integration as the software will naturally fit into the IT policy of the Institution. A *web-based interface* is essential, though the option of running the assessment on a *standalone* computer is a valuable feature. Reliable, efficient and immediate *technical support* for both educators and students is critical. *SCORM compliancy* is mandatory. *SCORM* (Shareable Content Object Reference Model) is a collection of standards and specifications for web-based e-learning. This facilitates the integration of the e-assessment system with the Learning Management System adopted at the institution (Carter et al., 2003; Maurice & Day; 2004; Pretorius et al., 2007).

3.2.2 Category 2: Question management criteria

3.2.2.1 Question types

This category describes the types of questions supported by the e-assessment tool. As discussed in Section 3.1.5, many different types of questions may be adopted (Fremont & Jones, 1994; Valenti et al., 2002). These include *Multiple Choice, Multiple Response, Fill-in-the blanks, Hotspot, Matching, Numeric,* *Calculated* and *Free Format (Short Answer)* questions. In addition, Pretorius et al. (2007) also include the *Information Screen* in this category. They state that it is important that the amount of text that can be input as information for the student is not restricted by a set number of characters, since the text included on this page is usually used to communicate a message to the student regarding the assessment.

Valenti et al. (2002) impress upon educators that when designing the types of question discussed in Sections 3.1.5 and this section, the structure of the questions must be aligned with the assessment strategy adopted for that module.

3.2.2.2 Question management criteria

In addition to offering varying question types, Pretorius et al. (2007) and other researchers discuss features that are necessary in an e-assessment tool to enhance the ease of creating questions electronically. These include:

- the *importing* of questions should be effortless if a text file, with a specified format, is used (Valenti et al., 2002),
- locating questions in a question bank should be facilitated by the automatic allocation of a *question code/unique identifier* to each question developed,
- *randomisation* of options is an essential feature to be included in the tool (Lewis & Sewell, 2007),
- *formatting* and the inclusion of *graphics* and *multimedia* elements should be achieved easily by the educator (Fremont & Jones, 1994; Carter et al., 2003).

In addition to regular scoring techniques, a facility for *partial and negative mark allocation* is an attractive feature to include. It should be possible to *preview questions* in the tool while they are being created. An additional feature that is useful is a *maths editor*, which facilitates the seamless inclusion of mathematical symbols into assessments.

3.2.3 Category 3: Test management criteria

Test Management criteria comprise a broad category. The three most important features included in this category are *categorisation* of questions, for easier searching and randomisation; allowing the *compilation of tests* with questions from each topic randomly displayed to students; and allowing the *printing* of tests and memorandums as required (Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002).

Control mechanisms (Fremont & Jones, 1994; Pretorius et al., 2007) are also required in the e-assessment tool, and these may include *time limits* for assessments, *randomisation* of distractors in a single question, a *limit* to the number of times the assessment can be taken by the student, *navigation* options that can be selected by the educator based on the test content and style, date and time of test *activation* that can be chosen, and students' *access* to the assessment for *revision* after completion.

Feedback options should allow the educator to decide whether the inclusion of feedback is required in the assessment in hand. The flexibility of choosing the format, and when and how the feedback is viewed, is also essential.

Analysis of student responses and test evaluation/analysis is another key element. This is achieved through well designed *reporting* features in the tool. These reports may include results and analysis of student performance using statistical analysis to present the difficulty level, discrimination index, and standard deviation of questions (Fremont & Jones, 1994; Pretorius et al., 2007).

Writing two decades ago, Fremont and Jones (1994) indicated that since the *Test Bank Database* is central to any e-assessment tool, it must support the creation and maintenance of both questions and tests. Specifically they outline the following requirements. A test bank must have the ability to:

- store a range of question types,
- include text, graphics and mathematical symbols,
- associate questions with informational fields, for example, question type, learning objective, cognitive level, level of difficulty, actual usage and statistics,
- integrate with other institutional software,
- be platform-independent,

- be adopted by various departments within the institution which will facilitate the sharing of expertise,
- be easy to learn and use,
- permit educators to select specific or random questions based on informational fields, when setting up an assessment, and
- analyse how students performed, either in a group or individually.

3.2.4 Category 4: Implementation criteria

For the successful implementation of an e-assessment system, it is important that the system is *stable, robust and fast* (Pretorius et al., 2007). Furthermore, it must incorporate high levels of *security* (Valenti et al., 2002) which include restricting access to assessments, window periods during which assessments are available, and limiting the number of logins and attempts that can be made by a student (Fremont & Jones, 1994; Pretorius et al., 2007). On-site *technical support* is another key feature, which will ensure a fast response time during problems (Fremont & Jones, 1994; Pretorius et al., 2007). The availability of *training* should ensure that educators utilise the system to its full capacity. *Service level agreements* with the developers of the system, are required (Pretorius et al., 2007).

3.2.5 Category 5: Interface criteria

Simply stated, the interface criteria category outlined by Valenti et al. (2002) states that the test environment should be *friendly,* should include a *graphical user interface* and should facilitate the *editing* of questions and tests easily.

Since the implementation of any form of e-assessment will be successful only if it is based on a sound, reliable and comprehensive system that satisfies all users' requirements, Parshall, Spray, Kalohn and Davey (2002) suggest that a comprehensive evaluation of comparable e-assessment tools should ensure that educators adopt the tool that best meets their requirements, as well as the needs of the students.

The criteria gleaned from the literature study, outlined in Table 3.1, were identified as important features that should be included in an e-assessment tool that is being considered for adoption. These identified features were used to create Framework Version 1a of this study, which is depicted in the top left of Figure 3.16. This will be expanded in Chapter 6, showing how this initial Framework, Version 1a, was used to facilitate the data collection.

3.2.6 Evolution of SEAT

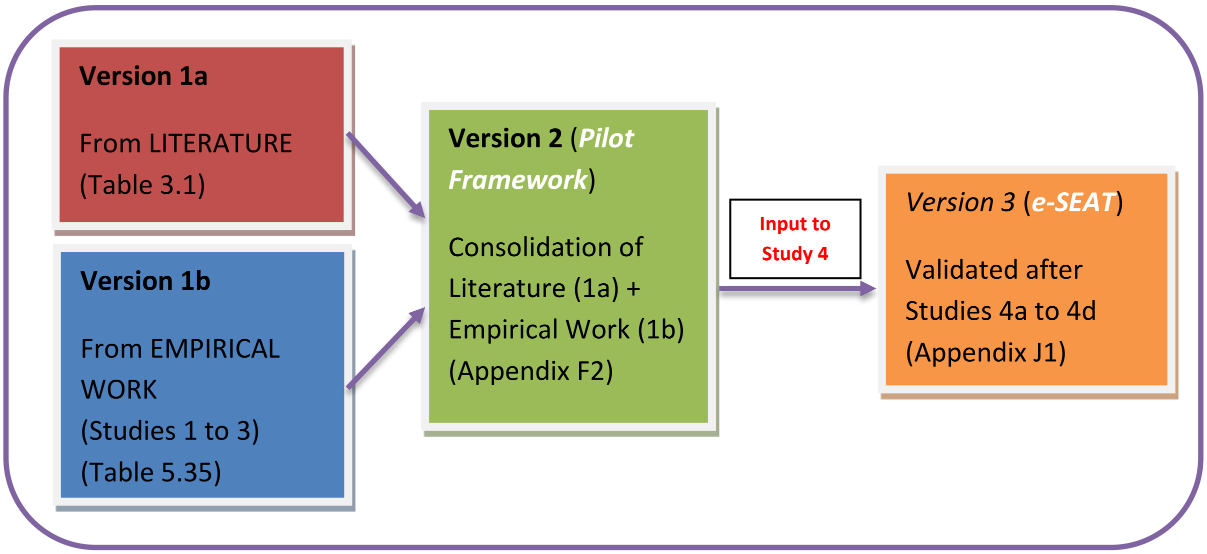


Figure 3.16: Evolution of the SEAT Framework

The initial framework synthesized from various literature sources is presented in Table 3.1. It is particularly geared to systems that administer MCQs. The framework presents eleven categories with 91 criteria in total, which the researcher deemed appropriate for inclusion in e-assessment systems. Although Sections 3.2.1 to 3.2.5 discuss only five categories, the researcher created eleven categories to ensure that the framework being developed is as comprehensive as possible. Additional criteria which were not included in Pretorius et al. (2007) nor in Valenti et al. (2002) were included from other literature sources. Some of these criteria were introduced for the first time in the framework in Table 3.1, and may not have been discussed in the literature studies. Furthermore, some criteria from the initial categorisations of Pretorius et al. (2007) and Valenti et al. (2002) were subdivided into smaller categories for more accurate classification.

The eleven categories created are presented in Table 3.1 and include *Interface Design, Question Editing, Assessment Strategy, Test/Response Analysis, Test Bank, Security, Compatibility, Import/Export, Ease of Use, Technical Support,* and *Training.* These eleven categories with 91 criteria contribute to the first version of the SEAT Framework (Pilot Framework in Appendix F1), which is presented in Chapter 6.

Table 3.1: Framework version 1a: e-assessment criteria derived from the literature

(Ref: Carter, Ala-Mutka, Fuller, Dick, English, Fone & Sheard (2003); Lewis & Sewell (2007); Maurice & Day (2004); Pretorius et al, (2007); Valenti, Cucchiarell, & Panti, (2002))

CATEGORY	CRITERIA
Interface Design	1. Intuitive to use (Carter et al., 2003) 2. Data must be accessible to users with special needs, for example, non-visual alternatives, font size, colour (Carter et al., 2003; Pretorius et al., 2007) 3. Can vary presentation of tests (Carter et al., 2003; Lewis & Sewell, 2007) 4. Email reminders of assessments due can be sent out (Maurice & Day, 2004)
Question Editing	1. Create the test and computerise personally (Carter et al., 2003; Lewis & Sewell, 2007; Pretorius et al., 2007; Valenti et al., 2002) 2. Authoring and Testing procedures must update records immediately and not at the end of the session (Lewis & Sewell, 2007; Maurice & Day, 2004; Valenti et al., 2002) 3. Can view and adapt other existing questions (Carter et al., 2003; Maurice & Day, 2004) 4. Questions can be imported and exported in non-proprietary inter-operable format (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007) 5. Ranges of parameters can be specified in questions (Lewis & Sewell, 2007; Pretorius et al., 2007) 6. Feedback can be provided for each question (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002) 7. Questions can be previewed both offline and online (Maurice & Day, 2004) 8. Revised questions are given global unique identifier (Maurice & Day, 2004; Valenti et al. 2002) 9. Questions which have not been answered can be deleted or amended (Lewis & Sewell, 2007), 10. Comments sent to question author directly (Carter et al., 2003, Maurice & Day, 2004) 11. Can allocate marks to questions and select a marking scheme (Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007) 12. Can combine questions into test (Lewis & Sewell, 2007; Pretorius et al., 2007)

	13. Can preview tests and see how they appear to students by testing them (Lewis & Sewell, 2007; Pretorius et al., 2007) 14. Authors should be able to approve/disapprove tests and add comments (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004) 15. Marks for each question and section are clearly displayed (Carter et al., 2003, Lewis & Sewell, 2007; Pretorius et al., 2007) 16. Variable parameters can be generated (Lewis & Sewell, 2007; Pretorius et al., 2007.) 17. Students can be forced to answer questions before moving on, if required (Pretorius et al., 2007) 18. Tests can be printed out (Carter et al., 2003; Lewis & Sewell, 2007) 19. Time taken for each question can be seen both for individuals and average student (Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007)
Assessment Strategy	1. Randomisation of questions and section order incorporated (Carter et al. , 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002) 2. Can incorporate branching of questions depending on users' responses (Maurice and Day, 2004; Pretorius et al., 2007) 3. Can display feedback and results as and if required (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002) 4. Can specify how many attempts can be made on a question (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Valenti et al., 2002)
Test and Response Analysis/Reports	1. Student access can be revoked while performance data is preserved (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007) 2. Groups can be set up and students added to a group (Pretorius et al., 2007) 3. Can view questions by metadata fields (Maurice & Day, 2004; Valenti et al., 2002) 4. Students have access to previous results, responses and markers' comments (Maurice and Day, 2004; Valenti et al., 2002) 5. Results can be accessed immediately (Lewis & Sewell, 2007; Pretorius et al., 2007) 6. Results can be accessed after a specific date (Carter et al., 2003; Lewis & Sewell, 2007; Pretorius et al., 2007) 7. Marks can be combined with marks from other tests (Carter et al., 2003; Pretorius et al., 2007) 8. Marks can be compared with others or group averages (Carter et al., 2003; Lewis & Sewell, 2007; Pretorius et al., 2007) 9. Access can be given to answers of other students (Lewis & Sewell, 2007) 10. Responses can be printed out (Maurice & Day, 2004; Valenti et al., 2002) 11. Markers' comments received through system or email to students (Carter et al., 2003)

	12. Late submissions by students of self-assessments prompt for the student to provide a reason for the late submission and warns the student of the penalty for such a situation (Maurice & Day, 2004; Valenti et al., 2002) 13. Can forward feedback to question where appropriate (Pretorius et al., 2007) 14. Markers emailed automatically if marking deadline is not met (Carter et al., 2003) 15. Analysis of mean and average score, discrimination and facility, frequency analysis (Carter et al., 2003, Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002) 16. Results tables can be ordered in various ways (Maurice & Day, 2004; Valenti et al., 2002) 17. Marks are displayable as percentages (Maurice & Day, 2004; Valenti et al., 2002.) 18. All attempts at a question can be viewed (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002) 19. Individual responses to questions can be viewed (Maurice & Day, 2004, Valenti et al., 2002) 20. Grades can be calculated over a series of tests (Carter et al., 2003; Pretorius et al., 2007) 21. Performance of different groups can be compared (Lewis & Sewell, 2007) 22. Performance in different subtopics (sections) can be compared (Lewis & Sewell, 2007) 23. Can view marks data without having access to names of students (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002) 24. Can correlate assessment data with other data such as age and gender (Carter et al., 2003, Pretorius et al., 2007)
Test Bank	1. Can draw random questions from a question bank (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002) 2. Students can create questions as their responses (Maurice & Day, 2004; Valenti et al., 2002)
Security	1. All tests and data are accessible to users who have explicit permission only, granted by access administrators (Pretorius et al., 2007; Maurice & Day, 2004; Valenti et al., 2002) 2. All data transmitted along the network is encrypted (Carter et al., 2003; Pretorius et al., 2007) 3. No material held on a server can be accessed by unauthorised persons (Carter et al., 2003; Pretorius et al., 2007) 4. Question authors must obtain permission of test author before altering a question (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002)

	5. Inability to amend or delete a test once taken by a student (Carter et al., 2003, Pretorius et al., 2007) 6. A global unique identifier is allocated automatically to tests (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004) 7. Ability to view entire tests for verification without the ability to change them (Maurice & Day, 2004) 8. Can restrict tests to particular IP addresses and domains (Lewis & Sewell, 2007; Pretorius et al., 2007) 9. Can modify results but must give a reason for the change (Maurice & Day, 2004) 10. All modifications and original marks are fully logged (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002)
Compatibility	1. Accessible from a standard, platform-independent web-browser, without additional plugins (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007) 2. Downgradable for users with early browsers (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004) 3. Customisable to provide a uniform interface with the rest of the institution's intranet, or learning environment (LE) (Carter et al., 2003; Lewis & Sewell, 2007; Pretorius et al., 2007) 4. Links seamlessly with other institutional systems so users can use their existing username and passwords (Carter et al., 2003; Lewis & Sewell, 2007; Pretorius et al., 2007)
Import/Export Data	1. Links seamlessly with other institutional systems so users can share student details and export marks directly (Carter et al., 2003; Lewis & Sewell, 2007; Pretorius et al., 2007)
Ease of Use	1. Short time to capture and set up (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002) 2. Little/no training required (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002) 3. Simple and fast login procedures required (Carter et al., 2003; Pretorius et al., 2007) 4. Intelligent help system – dependent on the user role and current activity (Maurice and Day, 2004; Valenti et al., 2002) 5. Speech synthesis for special need users (Carter et al., 2003) 6. Intuitive – no programming language to be learned (Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002). 7. Information about a group's tests can be specified (Lewis & Sewell, 2007). 8. Multimedia elements can be added with ease (Pretorius et al., 2007) 9. Students can return to the point where they left incomplete tests (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Valenti et al., 2002) 10. Students can enter symbols and foreign characters with ease (Maurice & Day, 2004; Valenti et al., 2002)

	11. Electronic certificate of test submission is sent to student (Lewis & Sewell, 2007; Pretorius et al., 2007)
Technical Support	1. Software easily installable with little effort and time (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007) 2. System works on Windows and UNIX servers (Carter et al., 2003; Maurice & Day, 2004) 3. Installation software easily available (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007) 4. Large numbers of concurrent users can be supported simultaneously (Pretorius et al. 2007) 5. Data storable in many formats – Oracle/Access or ODBC format (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al. , 2007) 6. Existing DB systems can be utilised (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007) 7. Possible to add, edit and remove user access administrators (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002) 8. Students and other users can be enrolled on and removed from the system (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al., 2007) 9. All students in a group can be removed from the system (Carter et al., 2003; Lewis & Sewell, 2007; Maurice & Day, 2004) 10. Access to details of all test purchases (Carter et al., 2003) 11. Sales and purchaser details can be transferred to separate e-commerce system (Carter et al., 2003)
Training	1. Little/no training, virtually self-learned (Carter et al., 2003, Lewis & Sewell 2007; Maurice & Day, 2004; Pretorius et al., 2007; Valenti et al., 2002)

3.3 Chapter conclusion

This chapter has presented a literature study in the use of MCQs in e-assessment. Section 3.1 defined MCQs, then benefits (Section 3.1.1) and drawbacks (Section 3.1.2) associated with their adoption were outlined. Thereafter a brief discussion was presented on the adoption of MCQs for HOTS (Section 3.1.3). The differing types of questions that can be created in an e-assessment system (Section 3.1.4) and finally examples of these various types were presented (Section 3.1.5).

Section 3.2 focused specifically on the five categories of criteria created by two leading researchers in the field of e-assessment (Pretorius, et al., 2007 and Valenti et al., 2002) that are used to evaluate e-assessment systems for the best fit. These include Technical Criteria (Section 3.2.1), Question Management Criteria (Section 3.2.2), Test Management Criteria (Section 3.2.3), Implementation Criteria (Section 3.2.4), and Interface Criteria (Section 3.2.5).

Finally, the initial framework developed by the researcher, based on the literature studies conducted in Chapters 2 and 3, was presented in Table 3.1. Its eleven categories are *Interface Design, Question Editing, Assessment Strategy, Test/Response Analysis, Test Bank, Security, Compatibility, Import/Export, Ease of Use, Technical Support,* and *Training.*

The main deliverable of this chapter is thus the initial SEAT Framework, garnered from the literature, and presented as Table 3.1. It answers the first part of Question 4, 'What does the literature suggest as appropriate requirements for electronic/online testing and assessment tools? Table 3.1 presents criteria that are appropriate for inclusion in a framework for evaluating e-assessment systems of the MCQ genre. It serves as the basis of the SEAT Framework and the criteria provide part of the foundation for Study 4, which involves a series of iterative action research studies conducted to further evolve the SEAT Framework. This is described in detail in Chapter 5, and more importantly, is the basis for the subsequent development of the e-SEAT Framework, discussed in Chapter 6.

CHAPTER 4 Research design and methodology

An introduction to the topic under investigation has been presented in Chapter 1, while Chapter 2 and Chapter 3 provide the findings of major literature reviews.

This chapter presents an overview of the research design and methodology adopted. Before discussing the approaches adopted for this study, a brief summary is presented of various types of research.

According to Merriam (2009), research is a systematic process by means of which the researcher learns about a phenomenon or an object, and knows more than he/she did before engaging in the research process. She further describes four ways in which researchers can engage in research processes by 'contributing to the knowledge base in a field (pure research), or improving the practice of a discipline (applied research), assessing the value of something (evaluative research), or addressing a particular problem (action research)' (Merriam, 2009:4). The present research primarily involves the fourth form — addressing a particular problem by action research, with the purpose of generating a framework to evaluate e-assessment systems of the MCQ genre. The long term consequences should also contribute to the second and third forms, in that application of the framework should improve the practice of e-assessment by providing a way to evaluate e-assessment systems.

This chapter commences with the discussion of the research foundations of this study in Section 4.1. This includes a presentation of Creswell's philosophical worldviews and the worldview of the present research (Sections 4.1.1 and 4.1.2). Thereafter the selected strategies of enquiry are outlined in Section 4.1.3, together with the strategies used in this study in Section 4.1.4. Finally, the role of research methods is briefly mentioned in Section 4.1.5. Research design and research methodology are covered in detail in Section 4.2 and Section 4.3 respectively. Section 4.4 elaborates briefly on the concept of a framework. The implementation of action research in this study is detailed in Section 4.5 and the six separate studies that make up the action research series are introduced. Finally the aspects of validity, reliability and triangulation are addressed in Section 4.6. The chapter conclusion is presented in Section 4.7.

4.1 Research foundations of this study

Figure 4.1 is based on Creswell (2009:5), who suggests that the research design for a study has three major components, namely the *philosophical worldview*, *selected strategies of enquiry* and the actual *research methods* used. Each of these is briefly explained in this section, and where appropriate, is applied to this study. The approaches used in this study are highlighted in red.

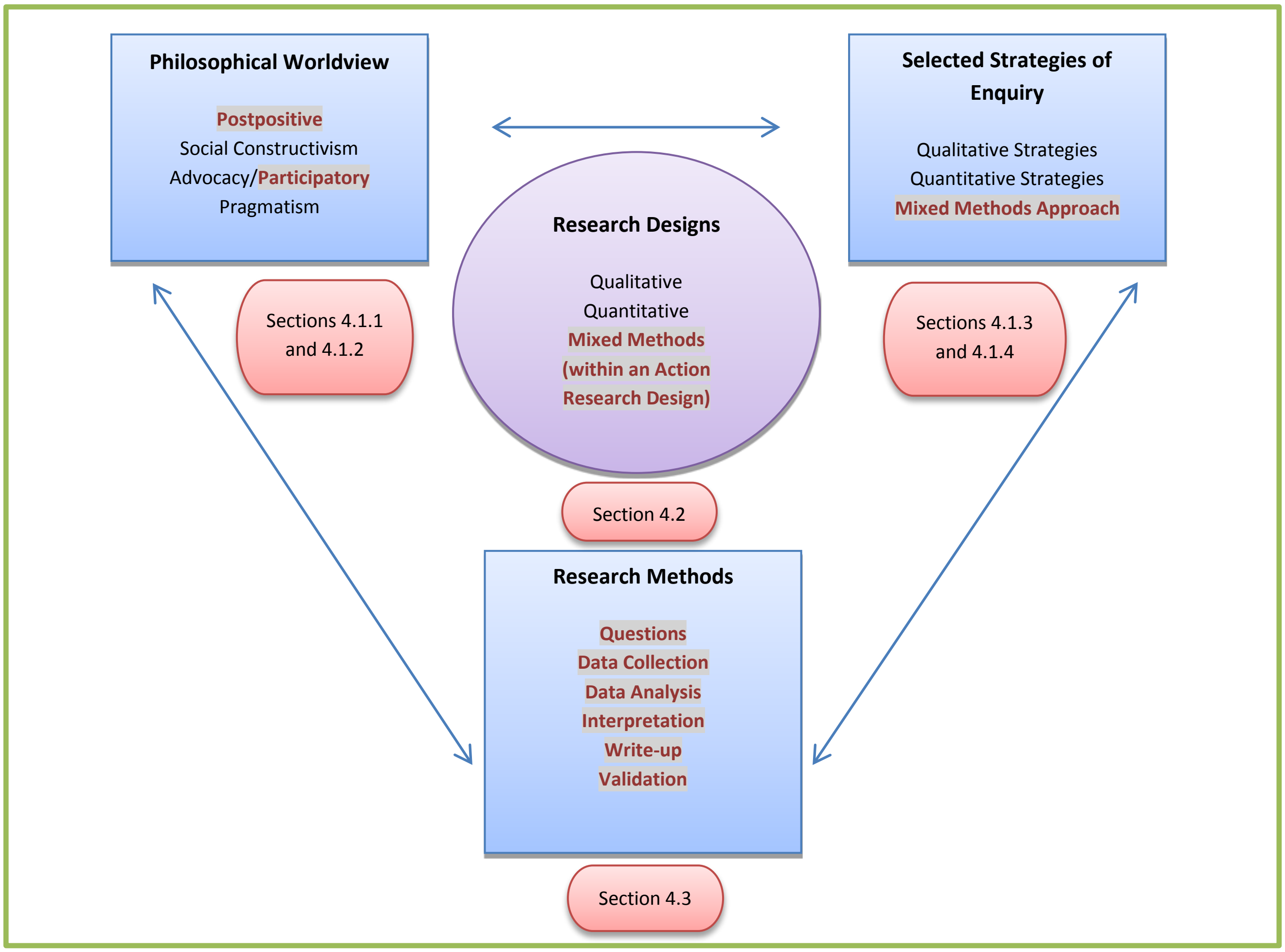


Figure 4.1: Creswell's framework for design:
The interconnection of worldview, strategies of enquiry and research methods
(Based on Cresswell, 2009:5)

4.1.1 Philosophical worldview

Creswell (2009) explains that the philosophical worldview refers to a general orientation regarding the world, relative to the nature of the research being conducted. These views are influenced by the discipline in which the research is being undertaken and by the researcher's prior experiences. Synonyms for this component include epistemologies and ontologies or research approaches. He suggests four different worldviews, namely, Postpositivism, Social Constructivism, Advocacy/Participatory, and Pragmatism, each of which is outlined in Table 4.1, followed by textual discussion of the respective views.

Table 4.1: Creswell's philosophical worldviews (synthesised by the researcher using Creswell, 2009:6-11)

Worldview	Description
Postpositivism (Section 4.1.1.1)	
Determination	Identifying causes of outcomes
Reductionism	Reducing ideas to a small set of concepts to test
Empirical observation and measurement	Measuring objective reality
Theory verification	Gathering data to support/refute a theory then revising it
Social Constructivism (Section 4.1.1.2)	
Understanding	Shaped by the researcher's experiences
Multiple participant meanings	Meanings constructed by participants' views
Social and historical construction	Based on culture, researchers understand the context of the research, through personal interaction with participants
Theory generation	Theories are generated from the data collected
Advocacy/Participation (Section 4.1.1.3)	
Political	Focuses on bringing changes in practice
Empowerment issue-oriented	Gives participants freedom to express themselves and has an agenda to change their lives
Collaborative	Creates debate and discussion to stimulate change
Change-oriented	Provides a voice for participants to engage as collaborative researchers in the study
Pragmatism (Section 4.1.1.4)	
Consequences of actions	Researchers are granted freedom of choice to choose methods and techniques to meet their needs, that is, they use what works
Problem-oriented	Focuses on a single problem and finding solutions to it
Pluralistic	Uses both quantitative and qualitative data to understand the research problem
Real-world practice-oriented	Research conducted in social, historical and political contexts, which reflect real-world practices

4.1.1.1 Postpositivist worldview

Postpositivism relates to the so-called 'scientific method' of conducting research. It is the traditional form of research and is founded on quantitative rather than qualitative studies. According to Cohen, Manion and Morrison (2011), the traditional scientific methodology of positivism is based on observation and experiment, and is restricted to what can be firmly established. It excludes efforts to gain knowledge by speculation and reasoning alone. Oates (2010) points out that the scientific method has two fundamental assumptions, namely, that phenomena in the world are ordered and predictable and, secondly, that they can be investigated objectively. The aim of the positivist approach is therefore to determine these universal laws and patterns. The basic techniques of positivism are reductionism, which involves decomposing complex concepts into smaller items that are more easily studied, and replicability, whereby an experiment will produce the same results each time it is repeated (Oates, 2010), and this verifies the underlying theory as shown in Table 4.1. This approach is particularly effective in the natural sciences (Cohen et al., 2011). For example, in physics and chemistry universal laws and formulae have been deduced that describe phenomena.

There is currently criticism of positivism (Cohen et al., 2011; Oates, 2010), particularly of its mechanistic approach that defines concepts in measurable terms, which can exclude personal experience, individuality, and matters of choice. Reductionism, though applicable in certain studies, is not always realistic; in many cases, it is more appropriate to study a phenomenon holistically and contextually. Similarly, repetition is not always possible – some studies can be conducted only once and would produce different results if investigated in other situations. Most importantly, for research such as the present one, the perceptions and interpretations of individuals are highly relevant in certain situations.

Creswell does not refer to positivism, but to postpositivism, which represents the thinking after positivism and poses a challenge to the positivist 'notion of the absolute truth of knowledge and recognising that we cannot be positive about our claims of knowledge when studying the behaviour and actions of humans' (Creswell, 2009: 7). Postpositivists aim to identify the causes of outcomes. Creswell's view of postpositivism, however, does include key features of positivism, such as the need to carefully observe and measure objective reality and, when studying human behaviour, to take numeric measurements of findings. Postpositivism also advocates beginning with a theory, then collecting data that either supports or refutes it, and making revisions before conducting further tests. Similarly, claims are made during the process of research and these claims are refined or abandoned as theory is tested.

Data and evidence are used to shape the developing knowledge. Furthermore, postpositivists seek to identify the causes of outcomes and assess these causes.

4.1.1.2 Social constructivist worldview

Social constructivism is the second approach presented in Table 4.1. It is often combined with interpretivism and is mainly adopted in qualitative research. Constructivism is based on the belief that individuals strive for an understanding of the world in which they live and work. Since this understanding emerges from subjective meanings of individuals' experiences, social constructivism relies on participants' views and personal interpretations of the subject being researched. This is achieved through the design of open-ended and general questions that facilitate discussions and interactions. Social constructivism highlights the importance of interaction with others, as well as culture and context in understanding events in society.

Researchers acknowledge that their own backgrounds also influence interpretation. Knowledge is created, based on this understanding and on participants' interpretations. Instead of commencing with a theory, researchers inductively generate a theory or pattern of meaning from the data collected (Alessi & Trollip, 2001; Creswell, 2009; de Villiers, 2012a).

4.1.1.3 Advocacy/Participatory worldview

In the 1980s and in the 1990s, a belief arose that the postpositivist paradigm was not appropriate for research relating to marginalised members of society or to matters of social justice. It was also felt that the constructivist approach did not address such issues. Therefore, the advocacy/participatory worldview arose as a basis for studies that combine research enquiry with political agendas. The goal of such research is reform that can contribute to improving the lives of participants and practices in the institutions where they live or work. Examples are studies on feminist perspectives, racial issues, critical theory and disability theory (Creswell, 2009).

The participative aspect of this worldview is that the researcher works practically and collaboratively with the subjects of the research, involving them as full participants and active collaborators. 'Participatory action is... focused on bringing about change in practices' (Creswell, 2009: 10). They may be involved in aspects such as assistance in designing questions, gathering data, and/or analysing information. The intention is that they will personally benefit from the outcomes of the research, since participatory research is focused on being an agent of change by actually implementing change in practice (Creswell, 2009).

The term, participatory, is often used in conjunction with action research, since action research involves its subjects/participants. Cohen et al. (2011) point out the participatory nature of action research – it is research by which participants contribute towards improving their personal practices, and also research in which the researcher frequently serves as a practitioner-researcher, investigating the evolution of his/her own product (de Villiers, 2012b).

Sections 4.2.1 and 4.2.3 of this chapter specifically address action research, which is the research design of choice for this PhD research and which is conducted in a highly participatory manner in the six studies in the action research series.

4.1.1.4 Pragmatic worldview

Unlike postpositivism, pragmatism stems from 'actions, situations, and consequences rather than antecedent conditions' (Creswell, 2009: 10). It is concerned with practical solutions to problems. Pragmatism can therefore be defined as an approach that evaluates theories or beliefs in terms of the success of their practical application. Pragmatists do not believe in an absolute unity or single truth, but rather view truth as what works in a situation at a particular time.

Pragmatism supports mixed methods research, and uses either quantitative or qualitative data or both – allowing the researcher freedom of choice with regard to the methods, techniques, and procedures that are most appropriate for their needs and purpose (Creswell, 2009). This approach offers the researcher the best understanding of the research problem. Thus pragmatism adopts 'multiple methods, different worldviews, and different assumptions, as well as different forms of data collection and analysis' (Creswell, 2009: 11-12).

4.1.2 Worldview of present research

Two worldviews are adopted in this research, namely, postpositivism and participatory, which are highlighted in Figure 4.1. Social constructivism also plays a role.

4.1.2.1 Postpositivism

Postpositivism is appropriate for the present research. Claims of knowledge were grounded in the behaviour and opinions of human participants, and careful measurements were undertaken when studying human behaviour and perceptions in the quantitative studies in this research. The first three studies, which investigate the nature and extent of adoption of e-assessment and multiple choice questions (MCQs) in Computing disciplines in South African Higher Education, generated data in the form of measurements of participants' personal usage of MCQs. The next set of studies, Studies 4, 5 and 6, obtained participants' opinions on the theoretical SEAT (Selecting and Evaluating an e-Assessment Tool) Framework, culminating in perceptions of, and experiences with, the electronic version of the framework, termed e-SEAT (electronically Selecting and Evaluating an e-Assessment Tool). These three studies involved the collection of numeric data from participants to support, refine or refute SEAT and, subsequently, e-SEAT. On this basis, revisions were made iteratively at each stage during the research process, before the next round of evaluation or testing was done.

4.1.2.2 Participatory

Creswell's third worldview has two aspects – the advocacy and participatory views. The political connotations of the former and its focus on marginalised members of society, are not relevant to this research. The participatory perspective, however, is highly relevant, because:

- the researcher worked collaboratively with the subjects of the research, namely users of MCQs, involving them as full participants in her studies,
- participants can personally benefit from the results of the research, as it contributes to improving practices in the institutions where they work, and
- a participatory approach is an integral part of action research, and action research (discussed in Sections 4.2.1 and 4.2.3) is used as the over-arching research design of this work.

Furthermore, the researcher was directly involved as the practitioner-researcher who designed the SEAT Framework.

4.1.2.3 Social Constructivism

Constructivism emphasises how theory and meanings are generated from participants' contextualised views and interpretations. Although the primary contribution of this research was to generate e-SEAT, an instrument for practical use, a theoretical framework also emerged from the interactions, that is, there is a social constructivist element. The researcher inductively generated a theoretical contribution on categories and criteria for evaluating systems that administer questions of the MCQ genre. This was based on the literature, on her own background and experience in e-assessment, which shaped her interpretation, and on interaction with human communities as data was collected from others in the field. In the questionnaires and interviews, the researcher included open-ended questions (as well as closed-ended questions), so that the participants could share their personal views. From these views, themes and patterns emerged.

To gain a deeper understanding, the researcher made personal visits to some of the participants, to appreciate the context and setting of MCQs and e-assessment and to see certain systems in use.

4.1.3 Selected strategies of enquiry

Creswell (2009:11) defines strategies of enquiry as the types of qualitative, quantitative, or mixed methods approaches that direct the research design of a study. Synonyms for this component include approaches to enquiry or research methodologies. These strategies have been summarised below.

4.1.3.1 Qualitative approaches

Qualitative research is an 'umbrella term that encompasses several philosophical or theoretical orientations' (Merriam, 2002: 15). Qualitative approaches refer to the use of non-numerical data, which include words, images, diagrams, and audio, generated from interview transcripts; researchers' notes; published and unpublished documents; memorandums; emails and faxes; and websites. These artefacts are gathered for research purposes during case studies, action research and ethnography.

Qualitative modes of analysis mainly involve textual analysis of verbal or written data. In general, qualitative data is used by interpretive and critical researchers (Oates, 2010). Interpretive theories aim at understanding the phenomenon being studied (Olivier, 2004). The researcher aims at understanding how participants 'make meaning of a situation or phenomenon' (Merriam, 2002: 6). Data can be collected through interviews and observations, and is inductively analysed to identify the underlying themes and patterns (Merriam, 2002). Interpretive studies are closely related to qualitative research.

Qualitative methods originated in the social sciences to facilitate the study of social and cultural phenomena (Myers, 1997), but are increasingly used in the applied sciences, such as Information Systems and e-Learning. According to Creswell (2009), qualitative research involves studying emerging questions and procedures; collecting data in participants' natural settings; and inductive analysis, building from the specific to the general, with reporting based on the researcher's interpretations of the meaning of the data. Qualitative research aims to understand how people 'interpret their experiences; construct their worlds and what meaning they attribute to their experiences' (Merriam, 2009: 5). Qualitative strategies and analysis techniques mentioned by Merriam and by Creswell include phenomenology, ethnography, grounded theory, textual analysis, critical qualitative research and qualitative case studies – each of which is defined in Figure 4.2.

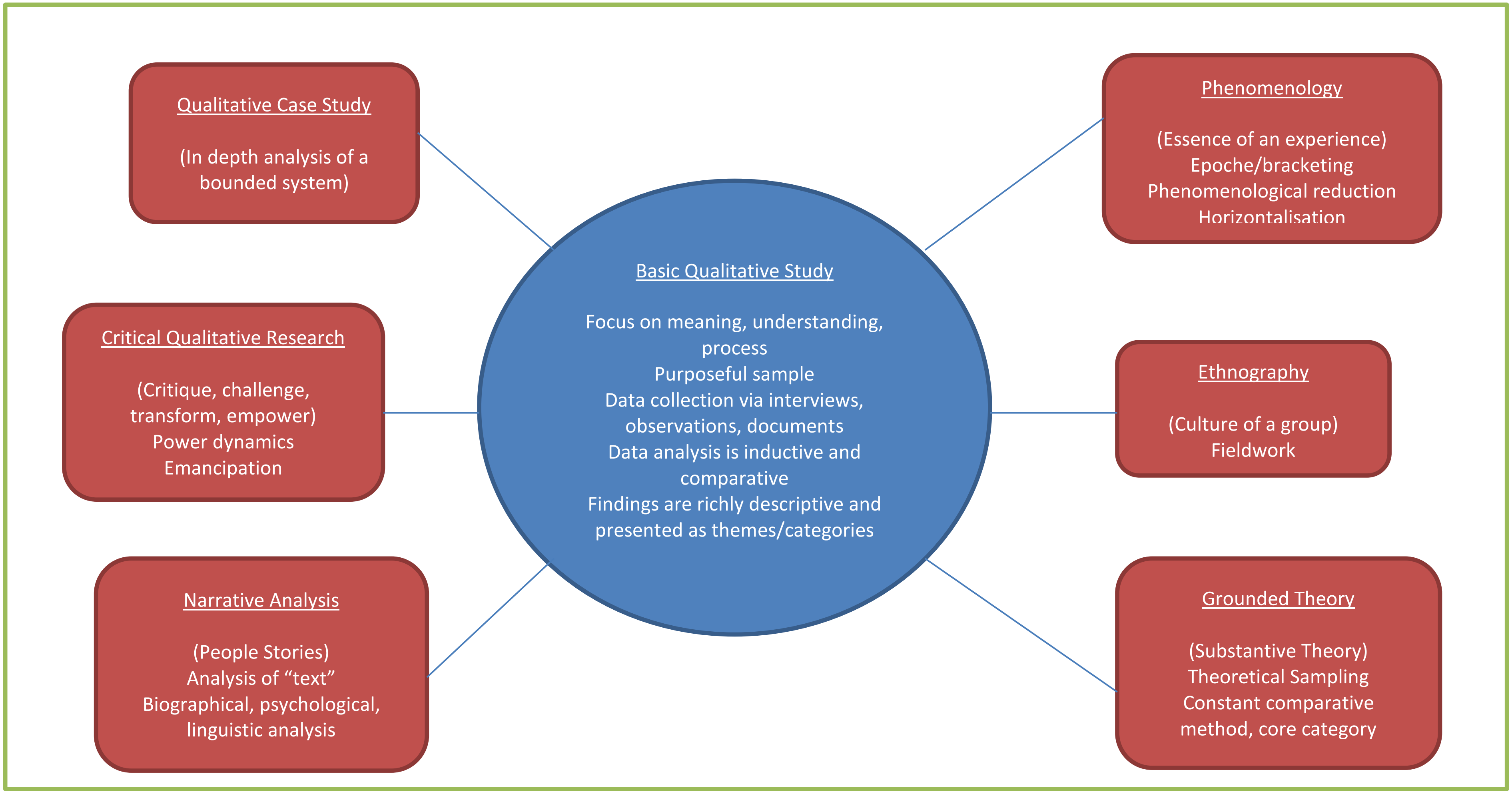


Figure 4.2: Types of qualitative research

(Merriam, 2009: 38)

4.1.3.2 Quantitative approaches

Quantitative research tests objective theories by examining the relationship between variables. (Creswell, 2009). Quantitative data refers to data based on numbers, mainly generated by experiments and surveys. Positivist and postpositivist researchers (see Section 4.1.1.1) are the main users of quantitative data and quantitative strategies of enquiry (Creswell, 2009).

With regard to analysing quantitative data, tables, charts and graphs are well-established techniques that allow the researcher and the readers to visualise the data patterns (Oates, 2010). These methods originated in the natural sciences for the study and measurement of natural phenomena (Myers, 1997). Quantitative data is frequently analysed by statistical analysis.

Quantitative approaches involve the type of numeric objective data already mentioned under postpositivism. Experimental designs are a classic form of quantitative study, while there are also non-experimental designs, such as surveys with closed-ended questions. Surveys can collect hard numerical facts, but they can also collect participants' opinions and quantify them by ordinal data in the form of Likert scaling. Both of these types of quantitative data were collected in the surveys in the present research. Certain quantitative strategies are illustrated on the right side of Figure 4.3.

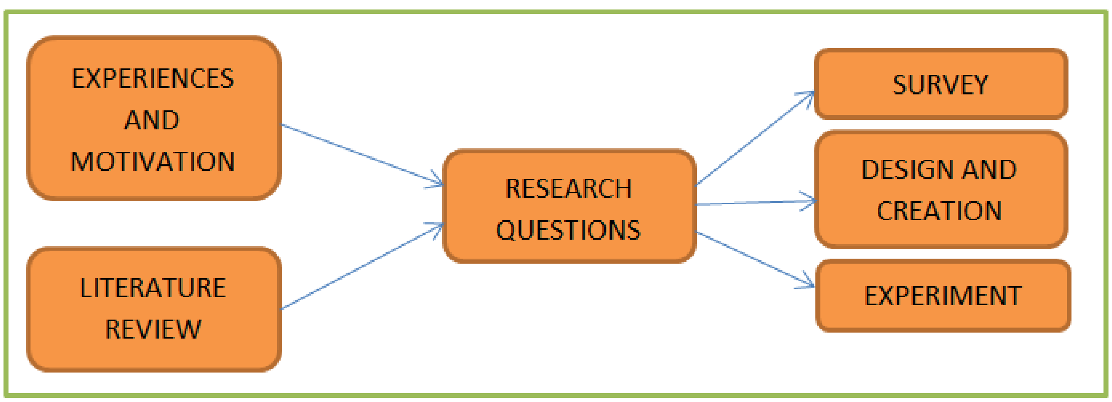


Figure 4.3: Types of quantitative research

(Adapted from Oates, 2010: 33)

4.1.3.3 Mixed method approaches

Mixed methods research involves combining different philosophical foundations and integrating quantitative and qualitative approaches in a study (Creswell, 2009). This helps to develop rich insights into areas of interest that are difficult to understand using solely qualitative or quantitative methods (Venkatesh, Brown & Bala, 2013). Cohen et al., (2011) propose that there should be a greater convergence between quantitative and qualitative approaches. Mixed methods are useful in uncovering information; providing dual perspectives and avoiding bias; confirming data; and helping researchers reach accurate conclusions (Cohen et al., 2011). This research approach is appropriate for real-world situations that are neither exclusively qualitative nor exclusively quantitative.

In mixed methods research, both quantitative and qualitative data is collected, analysed and interpreted in a single study or series of studies that relate to the same underlying phenomenon (Cohen et al., 2011). In studies termed 'concurrent mixed methods' (Creswell, 2009: 14), the researcher may converge or merge both quantitative and qualitative data. That is, both forms of data are collected simultaneously, and then the information is integrated when the overall results are interpreted. When 'sequential mixed methods' (Creswell, 2009: 14) are used, the findings of one method are expanded and elaborated by another method. Sometimes the process begins with an exploratory qualitative interview, followed with a quantitative survey method on a larger sample to determine whether the results can be generalised. On other occasions, the sequence is reversed, starting with a quantitative study and following it with qualitative research for elaboration (Creswell, 2009).

As outlined in Table 4.2, Creswell (2009) indicates that the research methods adopted in a study relate to the forms of data collection, analysis, and interpretation that are used. The table indicates how the methods vary between quantitative, mixed methods and qualitative strategies.

Table 4.2: Creswell's research methods (Creswell, 2009:15)

Quantitative Methods →	Mixed Methods	← Qualitative Methods
Predetermined	Both predetermined and emerging methods	Emerging methods
Instrument-based questions	Both open-ended and closed-ended questions	Open-ended questions
Performance data, attitude data, observational data, and census data	Multiple forms of data, drawing on all possibilities	Interview data, observation data, document data, and audio-visual data
Statistical analysis	Statistical analysis and text analysis	Text and image analysis
Statistical interpretation	Interpretation across databases	Themes, patterns or interpretation

4.1.4 Strategies adopted in present study

Based on the researcher's background and experience in e-assessment, the theory underlying her mixed methods research on the evaluation of e-assessment systems of the MCQ genre, was generated by social constructivism and participatory approach. The research methods used in her interaction with members of the academic community in the e-assessment domain, were questionnaires and interviews. These, together with personal visits to some participants, allowed the researcher to better understand the context and creation of MCQs and e-assessment.

Both quantitative and qualitative strategies were adopted in this study, i.e. it has used a mixed methods strategy of enquiry, as highlighted in Figure 4.1. In a sequential mixed methods style (Creswell, 2009: 14), the research commenced with a quantitative survey in Study 1 to determine the extent and nature of use of MCQs, followed by qualitative interviews in Study 2 to elaborate and gain more insight into this phenomenon. These studies were followed in turn by a series of four further studies, some quantitative and some qualitative. Each study had its own set of participants selected in a manner appropriate to that study. Specific details of each study – its data collection and analysis methods, and its sample of participants – are provided study by study in Section 4.5. However, this section overviews the different strategies used in the research as a whole.

Study 1 was a quantitative study that adopted a *survey-based research* strategy to gather background data into the extent and nature of usage of e-assessment tools and MCQs by South African Computing academics. This strategy was used to extract information from a limited number of respondents who represented a larger group of individuals, but were selected by the researcher, because they had the appropriate expertise and were willing and able to communicate about the information the researcher sought.

A combination of *quantitative structured questionnaires* and *qualitative semi-structured in-depth interviews* was the most commonly adopted strategy in the design of this research. This mixed methods design is ideally suited to gathering opinions, desires, attitudes and factual information (Hofstee, 2006: 122). Questionnaires focus on gathering similar data from a large group of respondents, in a standardised and systematic manner. Analysis involves using statistical analysis to find patterns in the data, in order to generalise to a larger population (Oates, 2010: 35). Survey research is usually quantitative in nature and aims to collect data from a representative sample of a larger population (Mouton, 2008). Surveys provide a 'numeric description of trends, attitudes, or opinions, of a population' (Creswell, 2009: 12) by focusing on a representative sample group.

The first form of qualitative strategy in this research was the acquisition of so-called secondary data, based on analysis of *literature reviews* to provide an overview of the scholarship in the field being studied. Findings of these literature studies provided secondary data, and presented perspectives and new categorisations of what has been researched by other authors. Chapters 2 and 3 present these literature studies.

Extensive literature reviews describe and show relationships between the various subcategories into which a field has been divided. This can identify possible gaps and thus help researchers within the field to generate new ideas. More importantly, new researchers to a field can benefit greatly from in-depth literature reviews, as it makes it easier for them to put the field into perspective based on the syntheses they develop (Hofstee, 2006). In-depth literature reviews can also present an outline of trends and debates, thus providing an overview of scholarship in a field of study (Mouton, 2008).

The literature reviews were used to obtain background knowledge into the types of e-assessment tools; trends in adoption of these tools; their potential benefits; and also constraints associated with the use of e-assessment tools. The findings from the literature, combined with data obtained in Study 1, contributed to the synthesis of knowledge about adoption patterns of e-assessment tools. This information was then used to further investigate the local usage of e-assessment tools in Study 2 and Study 3.

The second qualitative strategy was the collection of *qualitative data*. This involves analysis of data collected by primary researchers, either to verify an aspect of their research already presented or to answer a new research question (Hofstee, 2006). The researcher must be careful that any primary data being reused in secondary data analysis is reliable data (Hofstee, 2006). Qualitative analysis of primary data in this research served a dual purpose. The interview data collected in Study 2 verified the findings of Study 1; similarly, it confirmed the findings from the literature.

In this research, Studies 4c and 4d were solely qualitative. Participants' role in Study 4d, the Application Study, was to apply the e-SEAT Framework to evaluating an e-assessment system they are currently adopting. Their feedback on their interaction with the e-SEAT Framework served to record, the 'essence of their experience' (Figure 4.2). The analysis of text in Studies 4c and 4d used content and discourse analysis to extract themes from the interviews and obtained qualitative data to better understand participants' experiences.

Theory development, the third qualitative strategy, creates new ways of understanding aspects with which we are already familiar. This mainly involves testing an existing theory, or expanding its application. Sometimes, totally new theories may emerge. This approach may adopt modelling or philosophical reasoning (Hofstee, 2006). Such studies aim at developing new frameworks, models and theories, or refining existing theories or models, to explain a phenomenon (Mouton, 2008). In the present research, the intermediate outcome was not a theory as such, but a structured set of categories and criteria as a manual paper-based evaluation framework, named SEAT, for evaluating of e-assessment tools, while the final outcome was an electronic version of SEAT, named e-SEAT. The iterative development, evaluation and validation, first of SEAT, then e-SEAT, described in Studies 4, 5 and 6, was a practical and theoretical outcome of the final qualitative approach in this research, namely theory development.

4.1.5 Research methods

Research methods provide answers to the research questions posed in a study (Olivier, 2004). The overall design of this research, an action research design, is discussed in Section 4.2, while the set of research methods used for data collection and analysis, are explained in Section 4.3. Details of how of these methods were implemented, are tabulated and discussed in Section 4.5, which summarises the series of six action research studies, conducted in two phases.

4.2 Research design

This section overviews aspects of underlying research designs. An explicit research design is necessary in order to form a cohesive foundation for research. The section also introduces two possible designs for the present study, action research and design-based research, and then focuses on the research design chosen, namely action research.

A research design is the overall blueprint of how the researcher intends conducting the research (Miles & Huberman, 1994; Mouton, 2008). The research design adopted depends on three issues, essentially the:

- kind of research questions being investigated,
- purpose of the research, and
- research paradigms, principles and philosophies which underly them (Cohen et al., 2011).

According to Maree (2012: 70), a research design is a ‘plan or strategy which moves from the underlying philosophical assumptions to specifying the selection of respondents, the data gathering techniques to be used and the data analysis to be done’. Research that is focused on insight, discovery and understanding holds the greatest potential for making a difference to the participants (Merriam, 2009).

The design selected by the researcher should be based on ‘fitness for purpose’ (Cohen et al., 2011: 115), that is, the purpose of the research determines the methodology and design adopted. The research methods should be appropriate to answering the research questions being investigated. Thus, the researcher needs to select a research design that is in harmony with the research questions. Furthermore, to be successfully implemented, the researcher must be comfortable with the design selected, and must also understand the philosophical foundations underlying different types of research (Merriam, 2009). A wide range of research designs is currently available, so the researcher must select an approach that is highly appropriate for generating the kind of data required to answer the research question(s) (Maree, 2012).

Within the over-arching research design, there are three basic strategies of enquiry, according to Creswell (2009). These are *qualitative, quantitative and mixed methods* research, which were considered in Section 4.1.3. Qualitative methods adopt open-ended questions for data collection, and words for reporting, whereas quantitative methods use closed-ended questions for data collection and numbers for reporting. Quantitative designs were prominent in the late 19th century, until the middle of the 20th century. The use of qualitative designs increased in the latter half of the 20th century, followed by a transition to mixed-methods approach. Mixed methods research, as the name suggests, includes elements from both qualitative and quantitative approaches, often using both approaches simultaneously (Creswell, 2009). However, they can also be used sequentially for different studies in the same research project.

Mouton (2008) presents a simple analogy between an architectural design for a house and a research design for a research venture. This is shown in Figure 4.4.

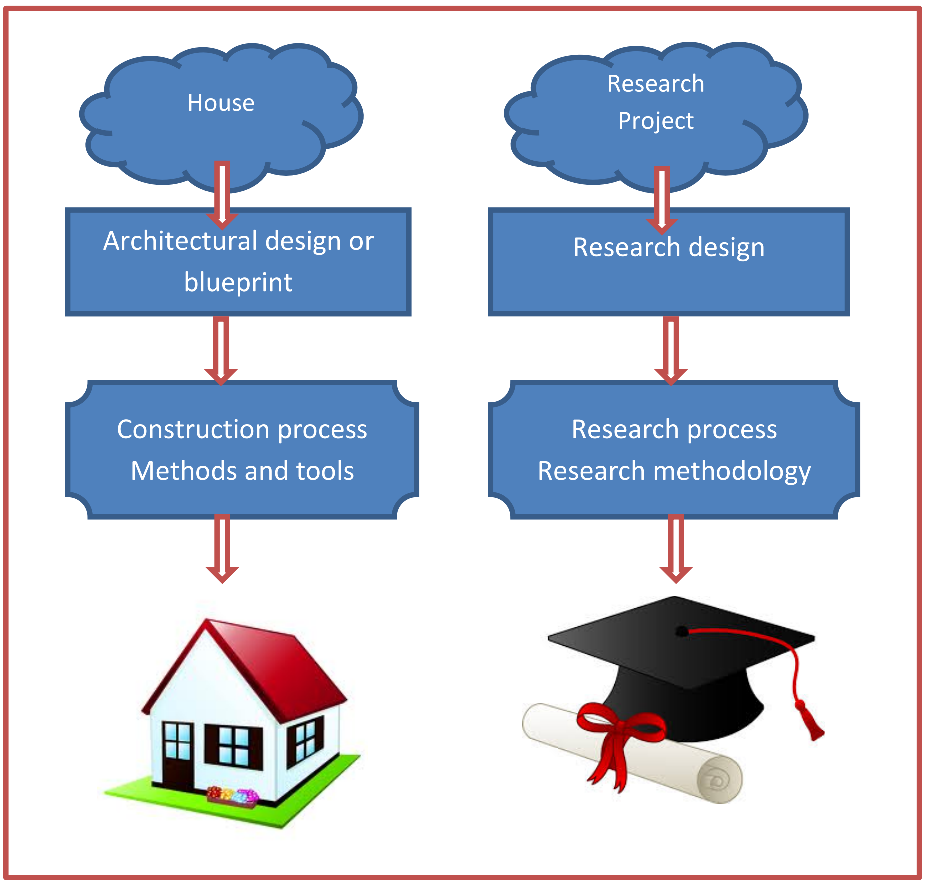


Figure 4.4: A metaphor for research design
(Adapted from Mouton 2008:56)

Mouton further presents a table (see Table 4.3), that serves as an appropriate follow-up to the issues raised by Cohen et al. (2011) and Merriam (2009) for selecting a research design, and to Figure 4.4. The table emphasises the relationships between the research design and the associated research methodologies, indicating how the product can be obtained by a process and by procedures, and how the research questions should be actualised by associated tasks.

Table 4.3: Mouton's research design and methodology (Based on Mouton, 2008:56-57)

Research Design	Research Methodology
Focuses on the end-product: What kind of study is being planned and what kind of results are required?	Focuses on the research process and the kind of tools and procedures to be used
Point of departure: Research problem or question	Points of departure are the specific tasks (data collection and sampling) to be undertaken
Focuses on the logic of the research: What kind of evidence is required to address the research question adequately?	Focuses on the individual (not linear) steps in the research process and on the most objective (unbiased) procedures to be employed

The next two sections review action research (Section 4.2.1) and design-based research (Section 4.2.2), followed by an explanation of the approach adopted in this study in Section 4.2.3.

4.2.1 Action research

Action research has been adopted in a variety of contexts, including community and development studies, classrooms, schools, universities, clinics, service providers, and information technology research. In the context of Information Systems research, Myers (1997: 248) defines action research as aiming to 'contribute both to the practical concerns of people in an immediate problematic situation, and to the goals of social science by joint collaboration within a mutually acceptable ethical framework'. The aim of action research is thus to solve current real-world, practical issues while simultaneously extending scientific knowledge (Baskerville & Myers, 2004). Action research is collaborative in nature, with the goal of adding to the body of knowledge already present in the field. Similar to case study research design, action research aims to gather information to inform a specific context or practice (Maree, 2012: 130).

Action research focuses on addressing a particular practical issue over a period of time, as it iteratively improves an intervention or generates a product. In many cases action research is used by professionals to assist them in investigating and improving their own personal working practices (Oates, 2010), as is the case in the present research.

Differing from ethnography, action research focuses on practical change. The researcher seeks to study the process and create organisational change, thus it is strongly oriented toward collaboration and change (Baskerville & Myers, 2004). As an agent of change, action research concentrates on practical solutions to concerns and complex problems in the real world, rather than working in a laboratory (Oates, 2010). Since the process is 'inherently transformative and developmental' the research aims to 'generate knowledge and action in support of liberating social change' (Maree, 2012: 124-5). Since the research may be extended to intervention, change may need to be facilitated within the group of people among whom the research is conducted.

With its collaborative ethos, action research actively involves participants in solving a problem or achieving an objective. In this participatory form of research, participants sometimes become co-researchers. In the context of computing, Baskerville and Myers (2004: 330) state that action research is a 'clinical method that puts Information Systems (IS) researchers in a helping role with practitioners'. The researcher collaborates with stakeholders who work in the situation under study, involving them as active participants (Oates, 2010) and using them to contribute actively to the generation of an intervention or product to alleviate the problem that was encountered (Maree, 2012). Involvement in planning, implementing, learning and evaluating process helps participants to contribute to the identification of the most effective way to achieve a goal (Hofstee, 2006).

The data gathering process in action research should be methodical, in order to gain clarity and insight into the issue being researched. Qualitative and quantitative methodologies can both be adopted in action research; however, it extends qualitative research through empowerment of participants. Qualitative methods provide deeper insight into the experiences and perceptions of research participants (Mouton, 2008). Furthermore, action research employs multiple methods to generate data (Oates, 2010).

Action research presents a method to explain why or why not, things work. It follows the principles of pragmatism which seeks to ask the right questions which would produce empirical answers to those questions (Baskerville & Myers, 2004).

The next two subsections present two different action research models, while the following two list respectively the benefits and challenges related to action research.

4.2.1.1 de Villiers representation

de Villers (2012a: 228) summarises action research by listing its five key features:

1. *Cyclic* – the stages are longitudinally iterative in nature, with each stage generating more knowledge.
2. *Participative* – participants collaborate with the researcher and are sometimes even called co-researchers in the study. Furthermore, the researcher is often a practitioner-researcher studying iterative versions of his/her own work, and working with participants in the process.
3. *Qualitative* – uses numbers in data collection, but data is usually more verbal.
4. *Reflective* – at the end of each cycle there is critical evaluation of the process which was observed, and the outcome of this introspection and reflection forms the basis of the actions in subsequent cycles.
5. *Responsive* – the study is flexible to adapt the intervention or product according to the findings of previous iterations.

The cycles, influenced by Zuber-Skerrit (1992), are represented in Figure 4.5. Note the position occupied by the researcher, indicating the central role that he or she plays.

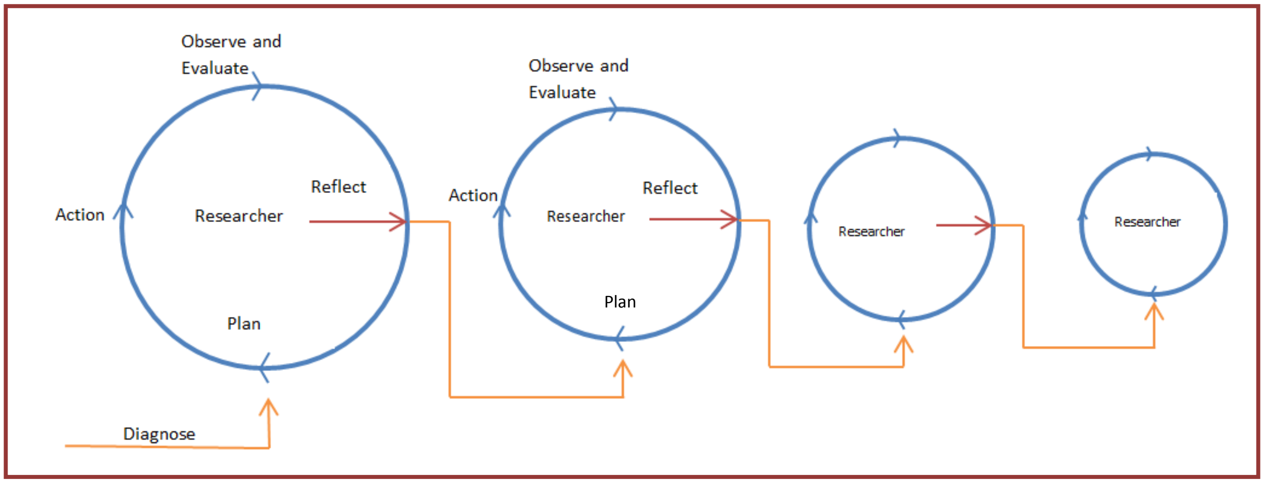


Figure 4.5: Action research model

(Adapted from de Villiers, 2012a)

The process of implementing action research involves the researcher planning an action to implement in the real world, actioning it, and then reflecting on the outcome, before planning the next iteration (Oates, 2010: 35). Thus action research is sometimes referred to as a four-stage cyclical process. The four stages which influenced the de Villiers model were outlined by Zuber-Skerrit (1992), namely: *Plan* – generating ideas; *Act* – testing; *Observe* – evaluating and generalising; and *Reflect* – understanding; then returning to Planning.

4.2.1.2 Maree's model

Similarly, Maree (2012) emphasises that the phases followed in action research are usually non-linear, since action research is iterative in nature. A continuous, bidirectional cycle, summarises the processes of action research as presented by Maree (2012), and depicted in Figure 4.6.

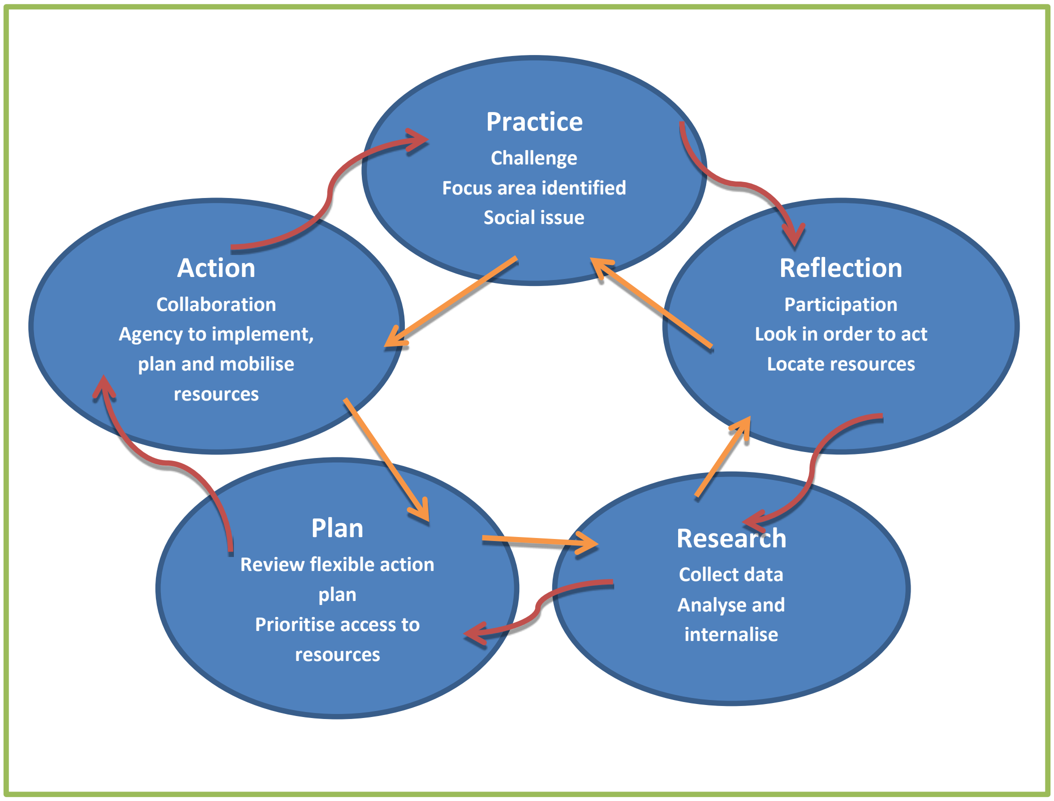


Figure 4.6: The cyclic process of action research
(Maree, 2012: 127)

This approach commences with identifying the problem, before proceeding to data collection using a variety of techniques, analysis of the data collected (*Research*), planning (*Plan*) and taking/implementing an action to resolve the problem (*Action*), and finally, assessing/evaluating the outcome of the action implemented (*Practice*) (Maree, 2012). The bi-directionality is an important feature (*Reflection*) by which the process can reverse itself and repeat the previous phase or phases.

Multiple data collection, data anlaysis, and evaluation methods may be adopted in action research, just as in a mixed methods design, although it is not an essential requirement (Maree, 2012).

4.2.1.3 Benefits of action research

Action research offers the following benefits:

- It addresses practical issues by feeding the results of the research back into 'practice' so that practitioners and organisations benefit directly (Creswell, 2009; Maree, 2012).
- Since participants are directly involved in the research, refusal to participate is usually low, as participants feel a sense of ownership of the results. Furthermore, direct participation allows respondents to present their viewpoints on opportunities and challenges they have faced (Maree, 2012).
- Both resources for data collection and time for respondent participation are cost-effective (Cohen et al., 2011; Maree, 2012).
- Since participants see that the research is focused on real-world problems, which address their interests, the relationship between the researcher and participants is strengthened, thus revitalising the learning community (Maree, 2012; Merriam, 2009).
- Action research promotes collaboration and empowers participants (Cohen et al., 2011).
- It leads to a solution of real-world problems and hence is interventionist in nature, meeting 'real' needs (Cohen et al., 2011).

4.2.1.4 Challenges associated with action research

During the implementation of action research, the researcher may face the following challenges:

- Earning the trust of participants can be a difficult task. If it is not achieved, the researcher runs the risk of not getting the correct insight into their perceptions and experiences (Maree, 2012). This trust can be established by acknowledging and respecting that all participants are different, and also differ from the researcher (Creswell, 2009). It is important that participants should be empowered, and not feel that there are power differences (Maree, 2012).
- Results should be verified by insiders, which may result in the findings being limited to the selected community and, often, not applicable beyond (Cohen et al., 2011; Maree, 2012).

Fundamentally, action research is designed to bridge the gap between research and practice, partly due to the failure of much research to positively impact on practice. It often serves as a means of empowering educators since it is a 'flexible, situationally responsive methodology that offers rigour, authenticity and voice' (Cohen et al., 2011: 361).

4.2.2 Design-based research

4.2.2.1 Design science and design research

Design science originated with Herbert Simon (Simon, 1981), who was a Nobel prize winner. He highlighted the difference between 'natural sciences' and 'design sciences'. Natural sciences are sciences such as physics, mathematics, and anatomy, etc. that describe and represent natural phenomena and relationships in the universe. In contrast, design sciences, which are also termed applied sciences, relate to man-made objects, such as those constructed in engineering, architecture, product design, information technology and education. Practitioners in design science engage in problem-solving processes, invention, and the creation of innovative products and interventions to solve authentic problems. Design science led to design research, where the main aim is to resolve real-world problems by generating and evaluating innovative artefacts to improve them. In the discipline of information systems, design research is known as 'design-science research' (DSR) and, in the domain of educational technology/e-learning, it is 'design-based research' (DBR) (de Villiers & Harpur, 2013). These two approaches are introduced in Sections 4.2.2.2 and 4.2.2.3 respectively.

4.2.2.2 Design-science research in Information Systems

The form of research termed design-science research (DSR) has its roots in engineering and in the design sciences as described by Herbert Simon (see Section 4.3.2.1). It is defined as a problem-solving activity which uses invention, intervention, evaluation and the measurement of impact (de Villiers, 2012b; Hevner, March, Park & Ram, 2004). Design-science research has been established as a research paradigm in the Information Systems discipline for a number of years, with researchers adopting it successfully, thus affirming its validity and importance (Offerman, Blom, Schonherr & Bub, 2010; Peffers, Tuuananen, Rothenberger and Chatterjee, 2008).

The most important complementary activities in generating IS artefacts in DSR, are 'build and evaluate', where the artefact can be a 'construct' (that is, a concept); a 'model' (or framework) in which constructs are combined; a 'method' involving steps to perform an activity; or an 'instantiation', which is an operational implementation (Hevner et al., 2004; March & Smith, 1995). Evaluation in DSR employs multiple evaluation methods, including observation, analytical and descriptive techniques, experiments, and testing (Hevner et al., 2004).

Despite design-science research being suitable for the development of business artefacts, Hevner and Chatterjee (2010) indicate that the adoption of this form of research has been slow in IS. Since IS managers are usually actively involved in the design activities of 'creation, deployment, evaluation and improvement of IT artefacts' (Hevner et al., 2004: 99), design-science research can prove useful in this context, with the challenge being the process of informing IS professionals of the strength and impacts of the new approach to research (Hevner et al., 2004; March & Smith, 1995).

4.2.2.3 Design-based research in Educational Technology

Design-based research is the educational technology form of design research. It is regarded as an extension of development research (de Villiers & Harpur, 2013). As evidenced by articles in educational technology journals, it is increasingly adopted for research into e-learning. Anderson and Shattuck (2012) did an overview of articles on design-based research that were published between 2002 and 2011. They identified 1940 publications, that is, almost two thousand, giving an indication of the extent to which design-based research is practiced. It blends empirical educational research with theory-driven design of learning environments, with the aim of building educational technology products as solutions to authentic teaching and learning problems (de Villiers & Harpur, 2013; The Design-Based Research Collective, 2003). Moreover, design-based research also focuses on developing theories, models and insights that are transferrable to other situations where emergent technology is developed for learning (Hay, Kim & Roy, 2005: 35). Figure 4.7 presents a design-based research model, highlighting its contextualized approach to solving complex problems, with an iterative ADDIE Model at its core. The ADDIE process involves cycling iteratively through the stages of **a**nalysis, **d**esign, **d**evelopment, **i**mplementation, and **e**valuation. The dual outcomes of design-based research are evident on the right.

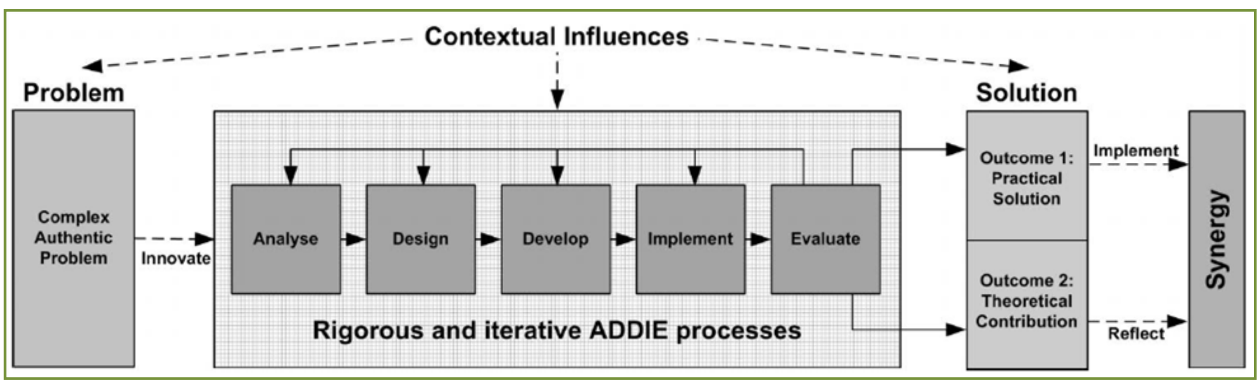


Figure 4.7: Design-based research model

(de Villiers and Harpur, 2013: 256)

As presented in Table 4.4, the products of design-based research are solutions to real-world problems. Insights emerge from contextualised studies, integrated with an iterative approach to problem solving. As Wang and Hannafin (2005: 5) succinctly state, design-based research advances design, research and practice simultaneously. Barab and Squire (2004) are of the opinion that design-based research should not be seen as a single approach, but rather a series of approaches that aim to produce both new solutions and new theories to impact on the educational process. Design-based research frequently adopts a mixed-methods approach to analyse and refine the interventions implemented (The Design-Based Research Collective, 2003). There may not be direct engagement with practitioners during the design process; however, practitioners 'reap the benefits of the research when it is completed' (Amiel & Reeves, 2008: 35).

A key feature of design-based research is that it produces dual outcomes, namely,

- a practical product to solve a real-world problem in an authentic setting, and
- a theoretical outcome in the generation of contextual and sharable design theories (Wang & Hannafin, 2005).

Table 4.4: Summary of features of design-based research model

(Adapted from de Villiers, 2012: 249)

Features of DBR Models
Addresses r*eal-world complex problems*
Solutions *grounded in existing theories* using technology as an aid
Solutions are *innovative* and novel, and produce interventionist technological support
The methodology adopted for studying the artefact is *systematic*
Design and evaluative processes are *iterative*
Theories and artefacts are *contextualized* in a particular setting
Empirical research is conducted on tangible, real-world items
Formative evaluation is used to *refine* the artefact
The *outputs* are real-world products that offer immediate value, as well as theories/constructs that are transferrable and customisable to other environments
Pragmatic theories are supported by evidence
Synergy of design and research
Rigorous and reflective testing of newly-designed learning environments

Experimentation can play a role in design-based research. 'Design experiments' with educational technology (Collins, Joseph & Bielaczyc, 2004) contributed to the early roots of DBR. In current DBR, prototypes are generated experimentally, and when a researcher claims that a design works in practice, such a claim must be based on evidence (Barab & Squires, 2004).

4.2.3 Research design selected: action research

In the disciplines of Computing and Information Systems, action research was disregarded for many years. However, there is currently an increased interest in this form of research. Since the 1980s, action research has been applied in IS and educational-technology systems. It has become a well-accepted form of research in e-learning studies (de Villiers, 2012a).

The researcher has adopted action research as the overarching research design for this PhD study, using mixed-methods strategies for the various studies in the series. The rationale for this choice is as follows:

- Action research and its features and processes were described in Section 4.2.1. It is an effective means of change and improvement and a powerful form of participatory research. Action research has a wide scope of applicability in terms of setting, the number of researchers involved, and the areas of study where it can be used (Cohen et al., 2011). Since it works best for educators on ‘problems that they have identified for themselves’ and assists peers since they 'can help each other in their professional development by working together’, it has been widely adopted by educators (Cohen et al., 2011: 344). Both these quoted factors are relevant to the present study, where the researcher identified the problem domain for reasons explained in the next bullet, and worked together with other educators as participants. These educators should benefit professionally from the long term outcomes of this work whereas design-based research is not essentially participative. Although action research originated in the social sciences, Baskerville and Myers (2004) believe that it provides an opportunity to make research in Computing domains more practically relevant, and they also point out that it is recognised as a valid research approach.

- In the initial step, the researcher should establish the purpose of the action (Baskerville & Myers, 2004). In the present research, the researcher identified a need for support related to the selection and adoption of e-assessment tools of the MCQ genre in Computing-related disciplines. Despite the increasing use of MCQs by South African Computing academics, there was no comprehensive instrument for evaluating and selecting an e-assessment system. As pointed out in Section 4.2.1, action research is often used by professionals wanting to improve their personal working practices (Oates, 2010). That was the case in this research, where the researcher was motivated to research this topic due to complexities she experienced in implementing MCQ tools. She therefore decided to investigate the phenomenon further.

- In the second phase, Baskerville and Myers (2004) suggest that there should be some practical action as an intervention to address the problem identified. The researcher acknowledged from personal experience and from the literature that there was a need for some form of framework to assist academics in evaluating and making decisions regarding which e-assessment tool to adopt. The design of the prototype e-SEAT formed the practical action in response to the identified problem.

- The participatory nature of action research considerably facilitated the process, since this study involved considerable participation from academic peers from 16 universities in South Africa. One hundred and eight (108) served as participants, as well as the international participants. Most of the participants were from Schools or Departments of Computer Science, Information Systems or Information Technology. This satisfied Baskerville's and Myers' (2004) step of ensuring that the reasoning and action are socially situated.

- Action research is iterative and longitudinal. By working together with practitioners and educators in the field of study for four years, an iterative research process was followed in the development and validation of a usable real-world framework to assist academics in the adoption or evaluation of e-assessment tools, based on their needs and requirements.

- The iterative nature of action research allows the researcher to determine the situation being studied and then make an intervention. Thereafter, the resulting situation is evaluated, and a further intervention made. This cycle of planning, acting, observing, reflecting and responding continues until the problem is resolved. It is important to note that the early cycles provide the researcher with opportunities for learning, as he or she uses the earlier attempts to improve the intervention in later cycles, till a solution is found (Olivier, 2004). In the present situation, the researcher embarked on fact-finding efforts in Phase 1 of the action research series, so as to establish the initial situation as a foundation for generating the framework.

- Since action research combines both action and research, it has become attractive to researchers and academics alike, as a powerful form of research (Baskerville and Myers, 2004; Cohen et al., 2011). It supported the present researcher in conducting a series of six studies, three of which helped establish the nature of the need and requirements for e-assessment tools (Phase 1), and three of which supported the study as the researcher progressively designed, developed and refined the instrument, which was called SEAT (Phase 2).
- Design research is currently being adopted in IS and educational technology as an underlying research design. A feature of design-based research, as applied in educational technology, is its characteristic dual output. Although the present research has a practical output in the form of e-SEAT, it does not explicitly generate theory. The evaluation framework can indeed be viewed as a contribution to theoretical knowledge in terms of the categories and criteria identified to judge the features of e-assessment systems. However, it is primarily an artefact for practical use, in line with the pragmatism of action research. This affirms the choice of action research rather than design-based research as the underlying research paradigm for this study.

4.3 Research methodology

Research methodology refers to the set of data collection and analysis methods used in conducting the research, and in processing and interpreting the findings. The research methods adopted by a researcher, refer to the instruments and procedures that are used to obtain and analyse the data (Cohen et al., 2011).

This section briefly overviews the general data collection and analysis methods used in this PhD research, but does not give details about the methods used in specific studies. That is done in Section 4.5 which provides concise details of the six studies in the action research process.

4.3.1 Data collection

This study adopts a mixed-methods research approach which combines quantitative and qualitative data.

Quantitative research reflects positivist or postpositivist philosophical assumptions. It collects numerical data which is typically analysed using statistics to examine relationships between variables (Creswell, 2009). In quantitative research, the researcher knows in advance what to look for. Therefore, the quantitative view is described as being 'realist' or 'positivist', due to its ability to uncover an existing reality (Oates, 2010).

When analysing qualitative data, researchers aim to find the 'meaning' that participants hold regarding the issue in hand, and usually do so inductively, bottom-up, to identify themes and patterns that emerge from the textual or verbal data (Creswell, 2009). A qualitative study is often adopted where there is a lack of theory in the field, or where the existing theory does not adequately explain a phenomenon. The product of a qualitative study is usually a description using words to present the findings of a phenomenon (Merriam, 2009). Merriam (2009: 14) presents four key characteristics of qualitative research: 'the focus is on process, understanding and meaning; the researcher is the primary instrument of data collection and analysis; the process is inductive; and the product is richly descriptive'.

Survey designs aim to obtain the facts and characteristics of a given phenomenon. When their results are presented in a numerical form, they are usually labelled quantitative (Merriam, 2009). Surveys gather information at a particular point in time with the aim of describing existing conditions or the relationship between events. They rely on large-scale data, which facilitates comparisons required for analysis (Cohen et al., 2011). In this study surveys – both questionnaires and interviews – were conducted to obtain quantitative and qualitative data.

Questionnaires are commonly adopted for gathering data, often numerical in nature, without the researcher's physical presence required (Cohen et al., 2011; Merriam, 2009). The larger the sample size, the more structured, closed and numerical the questions should be designed (Cohen et al., 2011; Oates, 2010). This makes the responses simpler to analyse. Open-ended questions can be included with smaller samples to obtain qualitative data from textual answers. Pilot testing is essential to ensure that the final questionnaire caters for all possible responses that can be easily envisaged (Cohen et al., 2011).

Interviews are regarded as the most flexible data collection tool, especially for complex issues, as they combine the use of verbal and non-verbal data collection techniques (Cohen et al., 2011; Oates, 2010). Since an interview is not an ordinary conversation, the researcher: has to capitalise on the time available for personal contact; should avoid interviewer bias; and be aware that anonymity may be a problem (Cohen et al., 2011). Interviews are not appropriate for large samples.

4.3.2 Data analysis

For quantitative data, *correlation-based research* can be used to generate correlation statistics to establish a relationship between two or more variables and to compare them, whether they are positive or negative correlations. *Descriptive statistical techniques* discover patterns in data, while complex statistical techniques help researchers to verify that the patterns they see in the data are accurate (Oates, 2010).

Content analysis examines the content of written documents. This method adopts both qualitative and quantitative approaches for data analysis. It is appropriate when an in-depth understanding of text is required, usually applicable to establishing ownership, or uncovering patterns and messages in textual data (Hofstee, 2006). Qualitative data analysis investigates the 'words, meanings, pictures, symbols, themes, or messages' in documents relevant to the study (Mouton, 2008: 165).

To analyse quantitative data, this study adopts quantitative content analysis as the data analysis technique. This is elaborated in Tables 4.5 to 4.10.

Qualitative research with its descriptive nature, can be enhanced by including 'quotes from documents, field notes, participant interviews, and excerpts from video tapes, electronic communication, or a combination of these' (Merriam, 2002: 5). These help to support the findings.

Quantitative data analysis can be done on qualitative data, but most qualitative analysis 'involves abstracting from the research data the verbal, visual or aural themes and patterns' that the researcher deems as relevant to the study (Oates, 2010: 267). According to Cohen et al. (2011), the creation of themes and patterns is also referred to as 'coding'. Coding is achieved by labelling each piece of text that identifies a specific thought or idea. This allows the researcher to find similar patterns among the textual data that has been collected. At times, a single piece of text may contribute to more than one theme and be assigned more than one code. Coding can be done manually or electronically.

4.4 Frameworks

Since this thesis relates to the development of a framework for evaluating e-assessment systems, this section briefly addresses frameworks and explains the nature of the framework developed for this research. First, the distinction is drawn between conceptual frameworks and evaluation frameworks.

A *conceptual framework* presents a collection of objects being studied – constructs and variables, and the relationships between them (Miles & Huberman, 1994). This is often depicted graphically. Leshem and Trafford (2007) consider the role of conceptual frameworks in the context of conceptualisation in doctoral research. They explain that the term, conceptual framework, is mainly used to describe a particular function and a set of interrelationships in a research process. Leshem and Trafford further state that conceptual frameworks give coherence to research by 'providing traceable connections between theoretical perspectives, research strategy and design, fieldwork and the conceptual significance of the evidence' (Leshem and Trafford, 2007:99). Based on her PhD research, Russell (2009) presents an evolving series of conceptual frameworks (also using the term 'systemic framework') to represent and analyse a complex adaptive system. Russell's tetrahedron-shaped framework represents processes, forms, contexts and materials, and models the adoption of educational technologies at various levels.

An *evaluation framework* is usually simpler, in a matrix-type format, and less focused on complex interrelationships. Macintosh and Whyte (2008) generated a coherent evaluation framework that used multiple methods and a variety of perspectives to evaluate e-participation initiatives in government. They proposed frameworks of criteria categorised under topics and converted criteria into evaluation questions. Yusof, Papazafeiropoulou, Paul and Stergioulas (2008) used an evaluation framework to evaluate health information systems (HIS). They found that traditional methods each evaluated a different aspect, therefore they developed a new approach built on previous evaluation models, namely a single integrated framework that examined technical, organisational and human factors in HIS's. They argue that evaluation is enhanced by combining various evaluation measures and classifying them into structured dimensions.

The purpose of this research was to develop a framework of categories and criteria to evaluate e-assessment systems of the MCQ genre. The work commenced with a study of existing frameworks and models, such as the literature sources discussed in Chapter 3 (Carter, Ala-Mutka, Fuller, Dick, English, Fone & Sheard, 2003; Lewis & Sewell, 2007; Maurice & Day, 2004; Pretorius et al, 2007; Valenti, Cucchiarell & Panti, 2002), which provide theoretical foundations for some aspects of e-assessment. However, none of these provide an ideal multi-facetted framework. Since they did not cover all aspects of e-assessment systems comprehensively, the researchers identified a niche to develop a comprehensive, multi-facetted evaluation framework for investigating e-assessment systems. The framework SEAT, presented in the present work, has elements of the previous frameworks, but it addresses the inadequacies in the identified literature on criteria essential for inclusion in e-assessment systems. Criteria were extracted from the literature, in particular from existing classifications. These categories of criteria were combined to synthesise the first version of a new structured framework that addresses a range of factors distinct to MCQ systems. SEAT was then extended through several more versions by the findings of empirical studies that elicited contributions and critique from experienced MCQ users. Such an evaluation framework can simplify the selection of a new e-assessment tool and facilitate comprehensive evaluation of existing e-assessment systems.

4.5 Implementation of action research in this study

The main purpose of action research is to allow researchers to longitudinally study aspects of practice with the aim of improving the practice. Likewise in this study, the researcher has reflected on the effectiveness of existing practice, the practice being e-assessment, with a view to improving the adoption of e-assessment tools. Since action research is about generating new knowledge based on enquiries conducted within specific practical contexts, this research, as depicted in Figure 4.8, was an iterative process focusing particularly on the practical context of e-assessment in South African tertiary institutions. In the course of the iterations, the research questions in Table 1.1 were addressed.

This work varies slightly from classic action research, which usually has a single aim, namely, to iteratively improve and evolve the intervention/product which is the output of the action research process. This research corresponds with classic action research in its longitudinal nature and sequence of cycles, but varies because Phase 1, comprising the first three studies, and Phase 2, comprising the next three, have different but related aims, as shown in Figure 4.8, which depicts the action research series.

Studies 1, 2 and 3 in Phase 1 cumulatively built a view of the e-assessment landscape in South African Computing education, with the objective of laying the foundation for the Evaluation Framework to be developed in Phase 2. Studies 1, 2 and 3 aimed to establish the nature, context and extent of adoption of e-assessment, particularly in Computing-related disciplines at South African Higher Education Institutions. In Studies 4, 5 and 6 in Phase 2, related to the requirements, creation, evaluation and application of the product of this action research, as well the impact of refinements, that is, Phase 2 is pure action research. The product of the action research was the Framework (SEAT and subsequently e-SEAT) for evaluating or adopting e-assessment systems and tools.
Phase 1 is presented in Chapter 5 and Phase 2 in Chapter 6.

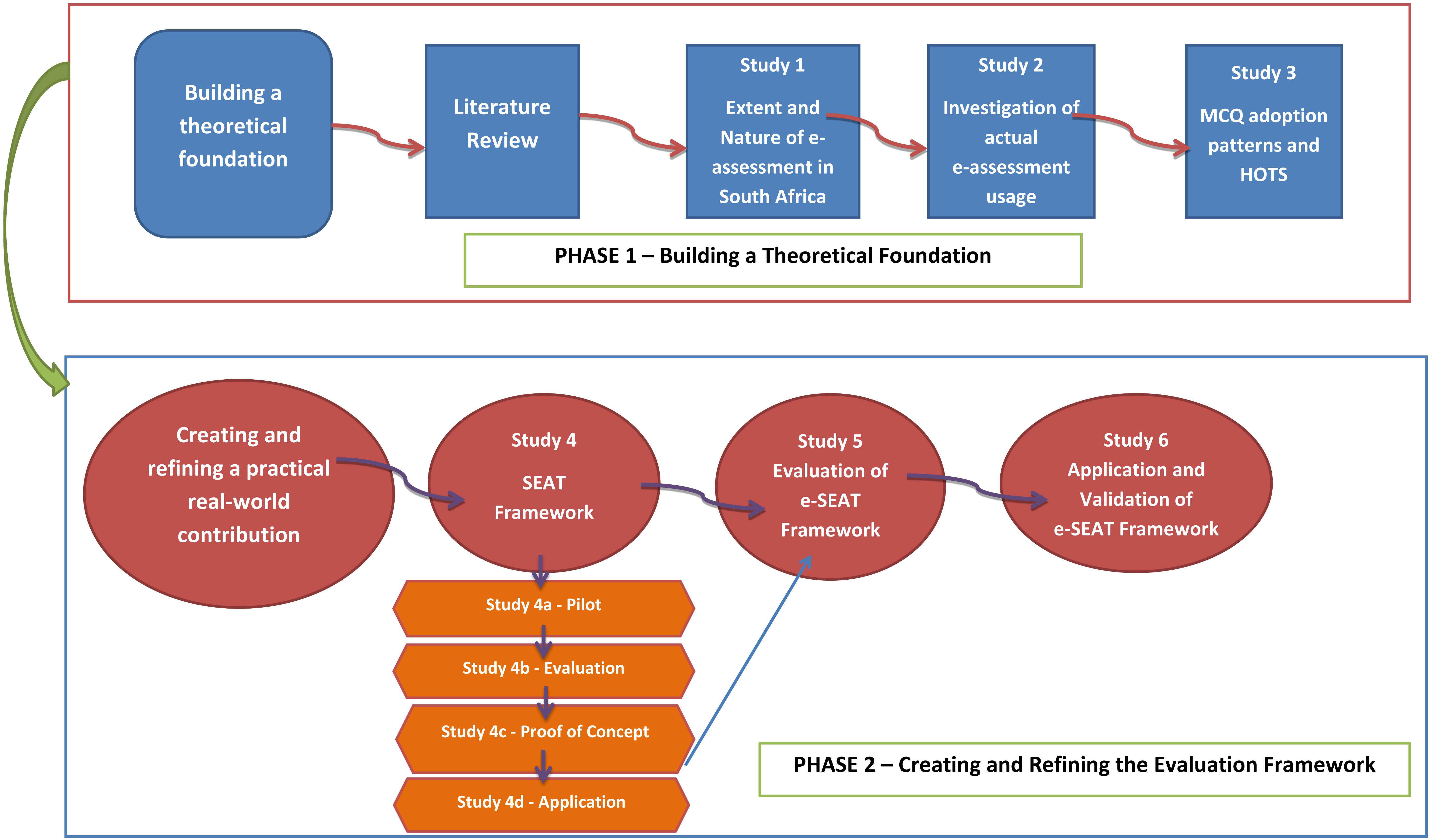


Figure 4.8: Action research applied in this study

PLEASE OPEN THIS PAGE TO VIEW A LARGE SIZE PRINTOUT OF **FIGURE 4.8,** WHICH ILLUSTRATES THE APPLICATION OF ACTION RESEARCH IN THIS STUDY

4.5.1 Study 1 – Extent and nature of usage

Research Question 1 asked "*What is the extent and nature of use of electronic assessment in Computing-related departments in South African universities?"* Study 1, summarised in Table 4.5, aimed at answering this research question.

Table 4.5: Summary of Study 1 as outlined in Figure 4.8

Study 1 (April - July 2009) - Extent and Nature of Usage	
Respondents	36 Participants from Computing-related departments at eight universities in South Africa
Data Collection	Survey Instrument – Questionnaire 1 (See Appendix C)
Data Analysis	Quantitative – basic statistical analysis Qualitative – content analysis (small component)
Purpose	The objective was to identify the nature and extent of usage of e-assessment at South African academic institutions, specifically within Computing disciplines (which include IS, IT and CS).

Phase 1 of this action research, which aimed at building a theoretical contribution, commenced with identifying the literature associated with the adoption of e-assessment tools both locally in South Africa as well as internationally. Little research was found that related to the use of e-assessment in South Africa. The researcher therefore conducted Study 1, with the objective of identifying the nature and extent of e-assessment usage at South African academic institutions, specifically within Computing disciplines (which include IS, IT and CS). This information was gathered through the use of a questionnaire directed generally at academics in Computing-related disciplines at all South African academic institutions. Convenience and volunteer sampling (Cohen et al., 2011) were used to acquire participants. Convenience sampling allowed the researcher to select participants for the Study from those nearest, or based on their availability. Volunteer sampling was useful for expansion of the participant population, because participants from a particular discipline, as well those who attended related conferences, were informed of the questionnaire and given the opportunity to participate voluntarily.

Study 1 was seen as a stepping stone to understanding the 'bigger picture' of e-assessment adoption at academic institutions in South Africa. Details on the sampling method used to acquire participants are given in Section 5.1 which describes this study and its findings in depth.

4.5.2 Study 2 – Actual e-assessment usage

To further investigate how e-assessment was adopted in South African tertiary institutions, Study 2, as presented in Table 4.6, targeted e-assessment users identified in Study 1 and referred by participants in Study 1, to gain further insight into adoption patterns of e-assessment. It contributed to answering Research Question 1.

Table 4.6: Summary of Study 2 as outlined in Figure 4.8

Study 2 (May 2010 – February 2011) – Actual e-Assessment Usage	
Respondents	72 respondents from both Computing and Non-Computing-related departments in 11 universities
Data Collection	Personal/Telephonic Interviews – Interview questions (See Appendix D) Informal observations, where possible
Data Analysis	Quantitative – basic statistical analysis Qualitative – content analysis
Purpose	To further understand actual usage of e-assessment at South African academic institutions and to identify respondents' opinions on criteria for evaluating e-assessment systems.

Based on the responses received from Study 1, Study 2 was conducted, directed again at the respondents from Study 1, who were categorised as 'users of e-assessment'. Study 2 took the form of interviews (personal and telephonic) with the targeted group of participants to gain further insight into adoption patterns of e-assessment and to obtain their opinions on features of e-assessment systems. These opinions contributed to additional criteria for the SEAT Evaluation Framework. Some participants also obliged during the interview by demonstrating the system/tool they had adopted. This informal observation was merely an enhancement to the interview and not a formal data collection method.

Since Study 1 had identified fewer users of e-assessment than anticipated from South African Computing academics, Study 2 was extended to non-Computing academics, through referrals that stemmed from the interviews (both personal and telephonic), in order to 'get a feel' of what was happening in the broader area of e-assessment. The group of participants was thus greatly extended by snowball sampling. This occurs when a small number of participants, who were carefully selected by the researcher, identify or refer the researcher to other possible participants, who have the characteristics required by the researcher. Snowball sampling is sometimes referred to as a 'chain referral method' (Cohen et al., 2011). This study is described in Section 5.2.

4.5.3 Study 3 – MCQ adoption patterns and HOTS

Research Question 2 asked "*What types of questions are being adopted in e-assessment systems in South Africa?* Thus Study 3, outlined in Table 4.7, studied the adoption patterns of the varying MCQ types available in e-assessment. Study 3 also utilized the same instrument, Survey 2, to answer Research Question 3, *"How appropriate are these questions (identified in Research Question 2) for testing higher order thinking skills (HOTS)?"*

Table 4.7: Summary of Study 3 as outlined in Figure 4.8

Study 3 (March – June 2011) – MCQ adoption patterns and HOTS	
Respondents	64 respondents from 15 South African Institutions (92 in total including international participants)
Data Collection	Survey Instrument – Questionnaire 2 (See Appendix E)
Data Analysis	Quantitative Analysis – basic statistical analysis
Purpose	To obtain information on the different types of MCQs adopted To understand how applicable these types of MCQs are to higher order thinking skills (HOTS).

From Study 2, it was identified that academics are making particular use of multiple choice questions (MCQs). Furthermore, it was noted that a few academics adopt these MCQs for more than just recall questions (lower cognitive levels of Bloom's taxonomy). Hence, Study 3 served a dual role – to obtain information on the different types of MCQs adopted, as well as how applicable these types are to stimulating higher order thinking skills (HOTS) in the students. This information was gathered through a short survey instrument distributed to the participants of Study 1 and Study 2. Study 3 thus ended Phase 1 of the study (see Figure 4.8), building a theoretical contribution regarding the nature and, as far as possible, the extent of use. Study 3 is discussed in Section 5.3 in Chapter 5.

The purpose of the next stage, Phase 2, as depicted in Figure 4.8, was to iteratively implement a practical 'solution' to the problems and issues identified in Phase 1, namely, the low adoption of e-assessment by Computing academics in South Africa. The solution envisaged by the researcher is a framework that academics can apply to evaluate e-assessment systems. Such evaluation would help them to select systems that would assist in their teaching and assessment processes. The development and refinement of this framework via action research spanned three studies, Studies 4, 5 and 6, of which Study 4 had four substudies (as shown in Figure 4.8), making six studies in all.

4.5.4 Study 4 – SEAT Framework

Study 4, as presented in Table 4.8, comprised four substudies in the action research series. Each substudy took the evolving SEAT Framework to a fresh set of participants, in efforts to continuously inspect and refine the framework from different perspectives and to formalise and incorporate criteria that are used in practice in South African higher education institutions for the selection and use of electronic/online testing and assessment tools. Study 4 thus answered both Research Question 4 "*What are the requirements for selecting or personally developing an electronic assessment tool?*" and Research Question 5 "*What categories and criteria should be incorporated in a prototype framework to evaluate electronic assessment systems?*"

Table 4.8: Summary of Study 4 as outlined in Figure 4.8

Study 4a (April 2012) – Pilot Study	
Respondents	2 Participants from UKZN
Data Collection	SEAT Pilot Instrument (See Appendix F2) and Questionnaire (See Appendix F3)
Data Analysis	Qualitative – content and discourse analysis
Purpose	To obtain initial critical feedback on the design, content and validity of the instrument to be used in the data collection process of the main Studies in this research
Study 4b (April – May 2012) – Evaluation Study	
Respondents	56 Participants from 16 Universities in South Africa
Data Collection	SEAT Instrument (See Appendix G)
Data Analysis	Quantitative – basic statistical analysis Qualitative – content analysis
Purpose	To determine which of the criteria identified in the Literature Reviewed in Phase 1, are essential for any e-assessment tool
Study 4c (May 2012) – Proof of Concept Study	
Respondents	3 expert users (UCT, UNISA and WITS)
Data Collection	SEAT Instrument (See Appendix H1) and Interviews (See Appendix H2)
Data Analysis	Qualitative – content and discourse analysis
Purpose	To gain insight into the criteria regarded as 'essential' for inclusion in the Framework
Study 4d (July 2012) – Application Study	
Respondents	7 expert users (UFS, UP, CPUT, UJ and NWU)
Data Collection	SEAT Instrument (See Appendix I), Questionnaire (see Appendix J2) and Follow-up Interviews, where required
Data Analysis	Qualitative – content and discourse analysis
Purpose	To apply the instrument developed (SEAT – Selecting and Evaluating an e-Assessment Tool) to an existing/adopted e-assessment system

Study 4, reported in Section 6.1 in Chapter 6, was an iterative study, commencing with Study 4a, where the initial SEAT Pilot Framework was developed on Survey Monkey and piloted in-depth by a sample of convenience, consisting of two Computing academics, who were colleagues of the researcher. SEAT is an acronym for Selecting and Evaluating an e-Assessment Tool.

The feedback received from the two pilot participants (structural, content and system related) was used to create Version 2 of the framework, named the SEAT Evaluation Framework, which was then distributed in Study 4b for evaluation by 'users' of e-assessment systems identified in Phase 1. Once again this version of the framework was developed on Survey Monkey. The aim this time however, was to determine which of the criteria identified in Phase 1, are essential for all e-assessment tools. Based on the responses received, non-essential criteria were removed. Further criteria, which were not previously identified by the researcher, but regarded as important by the respondents, were included in the framework. Finally, any significant comments made by the respondents were noted and the framework was adapted to create Version 3. Thus, Study 4b involved both quantitative and qualitative data analysis.

In Study 4c, Version 3, named the SEAT Proof of Concept Framework, was taken to three experts in the field for evaluation, that is, it was a purposive sample where participants were invited to take part in a proof of concept study. Post-evaluation telephonic interviews were conducted with each expert to gain insight into their comments, as well as to understand reasons for the low ratings provided to certain criteria that had been regarded as 'essential' by the respondents in Study 4b.

The final version of the SEAT tool was created after Study 4c. The wording of the framework was adapted to allow the 'application' study to take place as Study 4d. Seven expert users were selected from the respondent population of Phase 1. The application study required the seven respondents to apply Version 4, the SEAT Application Framework, to evaluate an existing e-assessment system that they used. The four versions and the associated studies are set out in Figure 6.2, while Studies 4a, 4b, 4c and 4d are respectively discussed in Sections 6.1.1, 6.1.2, 6.1.3 and 6.1.4.

This SEAT Application Framework served as a prototype for the ultimate tool to be developed, an electronic version named e-SEAT. The e-SEAT Framework incorporated automated scoring, calculation and reporting that provided category ratings and an overall rating to the e-assessment system being evaluated, as well as other features to support the user. e-SEAT was investigated in Studies 5 and 6.

4.5.5 Study 5 – Evaluation of e-SEAT Framework

Study 5, presented in Table 4.9, took the electronic framework developed (e-SEAT) to selected participants to evaluate the electronic version of SEAT. This study contributed to answering Research Question 6, "*How appropriate and effective is the proposed framework?*"

Table 4.9: Summary of Study 5 as outlined in Figure 4.8

Study 5 (Oct – Nov 2012) – e-SEAT Framework Evaluation	
Respondents	4 expert users (UKZN, DUT, UP and WITS)
Data Collection	Electronic Instrument e-SEAT (See Appendix J1), Questionnaire (See Appendix J3) and Follow-up Interviews, where required
Data Analysis	Quantitative – basic statistical analysis Qualitative- content and discourse analysis
Purpose	To evaluate the electronic version of SEAT (called e-SEAT – electronically Selecting and Evaluating an e-Assessment Tool)

With the feedback received from the Application Study in Study 4d, an electronic version of the SEAT instrument (named e-SEAT – electronically Selecting and Evaluating an e-Assessment Tool), was developed. The version was called the e-SEAT Evaluation Framework, since it was to be evaluated in Study 5. Study 5 involved four users of e-assessment, who evaluated the electronic version of the tool. These participants were users who had not taken part in any of the prior studies, and this provided a fresh and unbiased assessment of the tool created. The participants were therefore a purposive sample, invited by the researcher, and represented four tertiary institutions in South Africa. These participants gave positive feedback and only a few minor refinements were required following the evaluation.

Study 5 is discussed in Section 6.2 of Chapter 6.

4.5.6 Study 6 – Application and validation of e-SEAT Framework

The final study in the action research series was Study 6, in which the application and validation of e-SEAT were undertaken. Study 6 is summarised in Table 4.10. As was the case with Study 5, it served to answer Research Question 6, "*How appropriate and effective is the proposed framework?"*

Table 4.10: Summary of Study 6 as outlined in Figure 4.8

Study 6 (May – July 2013) – e-SEAT Framework Application and Validation	
Respondents	3 expert users (CPUT, MEDUNSA, UJ)
Data Collection	Electronic Instrument (See Appendix J1), Questionnaire (See Appendix J3) and Follow-up Interviews, where required
Data Analysis	Qualitative - content and discourse analysis
Purpose	To apply and validate the FINAL electronic framework developed (e-SEAT)

This version was called the e-SEAT Validation Framework, since it was to be validated in Study 6. A purposive sample of three experts in the field of e-assessment in South Africa, were asked to review and validate it, and apply it to evaluate an e-assessment tool that they use(d). By applying e-SEAT in this way, they also validated it by use. Thereafter they completed a short questionnaire on the applicability and usefulness of the e-SEAT Framework. In general they experienced positive interactions with e-SEAT, confirming that it was useful and intuitive to use. They made valuable suggestions for improvements, most of which were feasible, although some could not be implemented at this stage. The resultant and ultimate product of the action research was the e-SEAT Final Framework.

Study 6 is discussed in Section 6.3 of Chapter 6.

4.6 Validity, reliability and triangulation

This section defines the theoretical concepts of validity, reliability and triangulation. The concepts are revisited in Section 7.5 in Chapter 7, where it is outlined how each of the three concepts is implemented in this research.

4.6.1 Validity

Invalid research is of little value, hence validity is essential to effective quantitative and qualitative research (Cohen et al., 2011). Validity in research can be investigated on a high level by checking for the accuracy of the findings by using three factors. Researchers should ensure that:

- an appropriate process was adopted,
- the findings can be linked back to both the literature and the data, and
- the findings answer the research questions posed (Creswell, 2009; Oates, 2010).

The next paragraphs relate to validity on a more specific level.

Cohen et al (2011: 179) state that validity essentially demonstrates 'if a particular instrument in fact measures what it purports to measure'. Oates (2010) and Creswell (2009) address internal validity and external validity, mainly in the context of experiments, but Oates' discussion on external validity mentions threats from 'non-representativeness', which are equally applicable to the present research. The following aspects of non-representativeness should be avoided: too few participants, over-reliance on specific types of participants, and non-representative participants.

Table 4.11 below outlines the key features of validity in both quantitative and qualitative research.

Table 4.11: Validity in quantitative and qualitative research (Cohen et al, 2011: 182)

Bases of validity in Quantitative Research	Bases of validity in Qualitative Research
Controllability	Natural
Isolation, control and manipulation of required variables	Thick description and high detail on required or important aspects
Replicability	Uniqueness
Predictability	Emergence, unpredictability
Generalisabilty	Uniqueness
Context freedom	Context bounded
Fragmentation and atomisation of research	Wholism
Randomisation of samples	Purposive sampling/no sampling
Neutrality	Value ladenness of observations/double hermeneutics
Objectivity	Confirmability
Observability	Observability
Inference	Description, inference or explanation
Internal validity	Credibility
External validity	Transferability
Reliability	Dependability
Observations	Meanings

Qualitative data validity is addressed through 'honesty, depth, richness and scope of the data achieved' (Cohen et al, 2011: 179). Participant subjectivity, opinions, attitudes and perspective can at times present biases in qualitative data. To obtain qualitative validity, the researcher should apply procedures to check that the findings are accurate (Creswell, 2009). In this regard see Creswell's *qualitative validity strategies* at the end of this section. Quantitative data achieves validity through sampling, appropriate instrumentation and statistics, although there is usually a 'measure of standard error which is inbuilt' into the data (Cohen et al, 2011: 179). Quantitative validity must adhere to positivist principles. Cohen, Manion and Morrison stress that although varying interpretations exist of the definitions of validity in qualitative and quantitative research, they are not mutually exclusive.

Cohen et al (2011) also address validity in mixed-methods research, in which context it can be termed 'legitimation'. To achieve validity in mixed-methods studies, the following should be included: representation (using mainly words to capture past experiences and emerging situations); efforts to ensure that 'the results are dependable, credible, transferable, plausible, confirmable and trustworthy' (Cohen et al, 2011: 198); and integration of quantitative and qualitative methods.

To achieve validity in interviews, the amount of bias involved must be reduced. Moreover, for interviews to be valid, the inclusion of verbatim quotes from participants, reassures readers of *content validity* (Oates, 2010). To ensure that the questionnaires/surveys created, will generate data about the concepts being researched, *content validity* can be checked by ensuring that questions are brief, relevant, unambiguous, specific and objective – and thus a 'well balanced sample of the domain to be covered' (Oates, 2010: 227). Additionally, by correlating responses against each other, *construct validity*, can be attained, thus ensuring the researcher is 'measuring what the researcher thinks they are measuring through the questions posed' (Oates, 2010: 227).

Creswell (2009) recommends adopting multiple validity strategies to convince the reader of the accuracy of the findings. His eight primary validity strategies are:

- implementing triangulation – establishing themes based on the convergence of data,
- adopting member checking – requesting participants to check the accuracy of the data reported,
- using rich, thick description – providing multiple perspectives about a theme,
- clarifying bias – by self-reflection, showing how findings are shaped by the researcher's background,
- presenting negative information – that contradicts the theme discussed, making the data more realistic,
- spending prolonged time in the field – thus developing a deeper understanding of the phenomenon,
- using peer debriefing – to question the qualitative study, can present an alternate interpretation of the findings, and
- employing an external auditor – to provide an objective assessment of the research.

4.6.2 Reliability

The concept of reliability relates to whether the research approach is consistent (Creswell, 2009). For research to be reliable it must be dependable, consistent and replicable over time and across both participants and the instruments used (Cohen et al, 2011).

The focus is on repeatability; hence if a qualitative approach is reliable, it will be consistent across different researchers and various projects (Creswell, 2009). Furthermore, the research must prove that if it were to be replicated in a similar context, on a similar group of participants, then comparable results would be achieved. In qualitative research, reliability is achieved through 'credibility, neutrality, conformability, dependability, consistency, applicability, trustworthiness and transferability' (Cohen et al., 2011: 201).

In quantitative research, reliability is achieved through 'stability, equivalence and internal consistency' (Cohen et al., 2011: 200). Most quantitative research is positivist in nature and in the case of positivism, Oates (2010) points out that research instruments must be neutral (i.e. no leading questions), accurate and unambiguous. In the case of questionnaires, reliability measures whether the questionnaire would provide consistent results if administered to the same set of participants repeatedly (Cohen et al, 2011).

4.6.3 Triangulation

Triangulation is regarded as the use of multiple data generation methods to support and enhance the validity of research findings (Oates, 2010) and to converge various data sources or perspectives (Creswell, 2009). Types of triangulation that may be adopted, include (Cohen et al, 2011; Oates, 2010):

- Methodological triangulation – using different data collection methods on same object of study,
- Strategy triangulation – based on two or more research strategies,
- Time triangulation – use of cross-sectional or longitudinal time frames,
- Space triangulation – conducting research in more than one country or among more than one culture,
- Investigator triangulation – study undertaken by more than one researcher, and
- Theoretical triangulation (Oates, 2010) – research based on more than one theoretical perspective.

4.7 Chapter conclusion

This chapter discussed the research foundations of this study. Creswell's philosophical worldviews and those adopted in this study were presented in Sections 4.1.1 and 4.1.2. The selected strategies of enquiry were outlined in Section 4.1.3, together with the strategies used in this study in Section 4.1.4. Research methodology was discussed in Section 4.1.5. Thereafter research design and research methods were covered in Section 4.2 and Section 4.3 respectively. Section 4.4 briefly overviewed the concept of frameworks. The implementation of action research in this study was detailed in Section 4.5. Finally the aspects of validity, reliability and triangulation were highlighted in Section 4.6.

Following a brief summary of the various types of research available, this chapter presented an overview of the research design and methodology adopted in this study, namely, action research. This provided the theoretical background to action research, while the application of action research will be presented in Chapters 5 and 6, the chapters outlining the data presentation of the two phases in the series of action research conducted. Some of the concepts mentioned in the overviews of validity, reliability and triangulation will be reviewed in Chapter 7 to indicate how they were implemented in this research.

The main deliverables of this chapter are:

- the motivated choice of action research as the overarching research design, and
- a map of all the studies in the action research series that will develop, evaluate, refine and validate the emerging framework. This sets the scene for the empirical research that follows.

CHAPTER 5 Data presentation and analysis of Phase 1 Studies

The previous chapter on the research design and methodology of this PhD study outlined the six studies of the action research series and depicted them graphically in Figure 4.8. This chapter focuses on the findings of Studies 1 to 3 which make up Phase 1 of the series. In an attempt to partially answer Research Question 4, these three studies set out to establish the nature, context and extent of adoption of e-assessment, particularly within Computing-related disciplines at South African Higher Education Institutions.

Sections 5.1, 5.2 and 5.3 discuss Studies 1, 2 and 3 respectively. A description of the evolution of the SEAT Framework is presented in Section 5.4, followed by a summary of the chapter in Section 5.5.

Figure 5.1, which is the first section in Figure 4.8, illustrates the progression of Studies 1, 2 and 3. These studies set the context for research on requirements for, and evaluation of, e-assessment systems and tools. In this way they build a theoretical and conceptual basis that makes an important contribution to the foundation of the Evaluation Framework (SEAT) which is created and refined in Phase 2 and converted to the electronic version, e-SEAT.

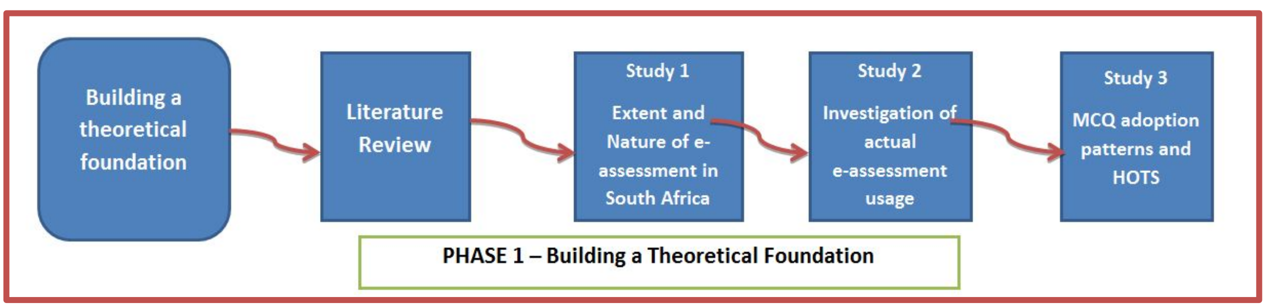


Figure 5.1: Phase 1 of the action research series

5.1 Study 1

The time-independent and location-independent nature of Information and Communication Technology (ICT) can provide ‘decongestion of overcrowded education facilities, support for students and educators, and a valuable opportunity for specific groups of students, if the learning material is accessible to them’ (Ardito et al., 2006: 12). This statement applies not only to instruction and learning via e-learning, but is also relevant to assessment supplemented by ICT technologies. Many universities internationally use e-assessment for at least some portion of their assessment programs. Research has indicated that many tools and systems exist that can create, deliver, score, report and analyse both summative and formative assessments, and provide various other customised online and paper-based testing and reporting services (Harrington & Reasons, 2005; JISC, 2007).

The use of e-assessment and computer-based testing is on the increase in South Africa. In Study 1, conducted in 2009, and outlined in Figure 5.1, a survey was conducted to investigate the extent and nature of use of e-assessment tools in Computer Science (CS), Information Systems (IS), and Information Technology (IT) academic units in South African tertiary institutions, as well as satisfaction on the part of academics who are users. This set the context for the further studies in Phase 2 of this research.

Some of the material in Section 5.1 is based on a conference paper presented at the South African Computer Lecturers’ Association (SACLA) Conference in 2010 (Singh & de Villiers, 2010). The research presented in the publication, was conducted by the researcher as an integral part of her PhD studies and was also used for the conference paper.

The findings include both quantitative (Section 5.1.5) and qualitative aspects (Sections 5.1.6 and 5.1.7), which are presented separately for data obtained from existing users of e-assessment (Sections 5.1.5 and 5.1.6) and data from non-users (Section 5.1.7) who are potential users.

5.1.1 Introduction to Study 1

This study, as summarised in Table 5.1 below (similar to Table 4.5 in Section 4.5.1), aimed to establish the context of adoption of e-assessment tools within Computing-related academic departments/schools at South African tertiary institutions.

Research Question 1	*What is the extent and nature of use of electronic assessment in Computing-related departments at South African universities?*

Study 1 also used open-ended questions to further investigate academics' satisfaction with the use of these tools.

Table 5.1: Summary of Study 1 as outlined in Figure 5.1

Study 1 (April - July 2009) - Extent and Nature of Usage	
Respondents	36 Participants from Computing-related departments at eight universities in South Africa
Data Collection	Survey Instrument – Questionnaire 1 (See Appendix C)
Data Analysis	Quantitative – basic statistical analysis Qualitative – content analysis (small component)
Purpose	The objective was to identify the nature and extent of usage of e-assessment at South African academic institutions, specifically within Computing disciplines (which include IS, IT and CS).

The extent of South African usage was found to be low, but on the increase. There were 36 respondents from eight institutions, sixteen of whom are regular users of e-assessment, mainly using multiple choice questions (MCQs). The systems were employed more for formative than for summative assessment. Most usage was for large first-level classes. The benefits (Table 5.9) and disadvantages/barriers (Table 5.10) mentioned by respondents correspond well with those identified in the literature study (Sections 2.5.7 and 2.5.8).

5.1.2 Background to Study 1

During a seven year period as first-level Information Systems coordinator at the University of KwaZulu-Natal, the researcher faced a major challenge in efforts to implement a software tool for e-assessment. Due to the large number of students, approximately 1600, distributed over two campuses, the primary method of assessment adopted in the School of Information Systems and Technology (IS&T) for its entry-level students was paper-based multiple choice questions. The large administrative and marking load associated with these forms of assessment prompted the researcher to investigate various e-assessment tools for the judgment of MCQs, including: SAM (Skills Assessment Manager by Pearson), Hot Potatoes, ExamView, EzTests and CourseCompass. During the testing of these software tools for implementation purposes, various problems were encountered, two of the major issues being:

- tool interfaces were not easy for students to understand, and
- the administration associated with implementing the tools was laborious, hence did not provide motivation for the academic administrator to implement these technologies.

Following this failure to implement an e-assessment application in the School of IS&T at the University of KwaZulu-Natal, yet taking cognisance of recent advances in educational technologies, the researcher set out to investigate the current level of usage of e-assessment tools within Computing-related academic departments at South African tertiary institutions, and to determine the levels of satisfaction or dissatisfaction of academics using these technologies. To this end, research was undertaken in IS, IT and CS departments, aiming to determine the extent and nature of use of e-assessment:

Research Question 1 is quantitative in nature, while the follow-on regarding satisfaction required qualitative answers. These research questions were developed due to:

- the researcher's personal interest in the area of study,
- motivation for the knowledge outcomes to be achieved, and
- suggestions from the literature of areas where further research is required (Cheng, Jordan & Schallert, 2013; Christakoudis, Androulakis & Zagouras, 2011; Deutsch, Herrmann, Frese & Sandholzer, 2012), as was explained in Section 1.2.1.

5.1.3 Participation: by faculty and department

The questionnaire, which is Questionnaire 1 in Appendix C, investigated aspects such as participants' background details. These included the *Institution, Faculty, School, Department/Section* they belonged to as well as the *Position* they held.

Questionnaires were e-mailed to the IS, IT and CS academic departments at the nine higher education institutions where Gatekeeper Consent had been obtained, with a request to the School or Department head to inform staff of the research being undertaken, and hence request them to participate as volunteers. Since very few respondents participated to the initial request, the researcher then obtained permission from the Heads of School/Department to email staff directly. In an effort to attract more participants, the local mailing lists of the interest groups for Computing-related academics, namely South African Computer Lecturers' Association (SACLA) and South African Institute of Computer Scientists and Information Technologists (SAICSIT), were also used. These collective contact methods resulted in 36 participants, comprising a sample of volunteers.

The questionnaire commenced with a question on the participants' personal profile as well as to ascertain in what type of school/department he/she worked.

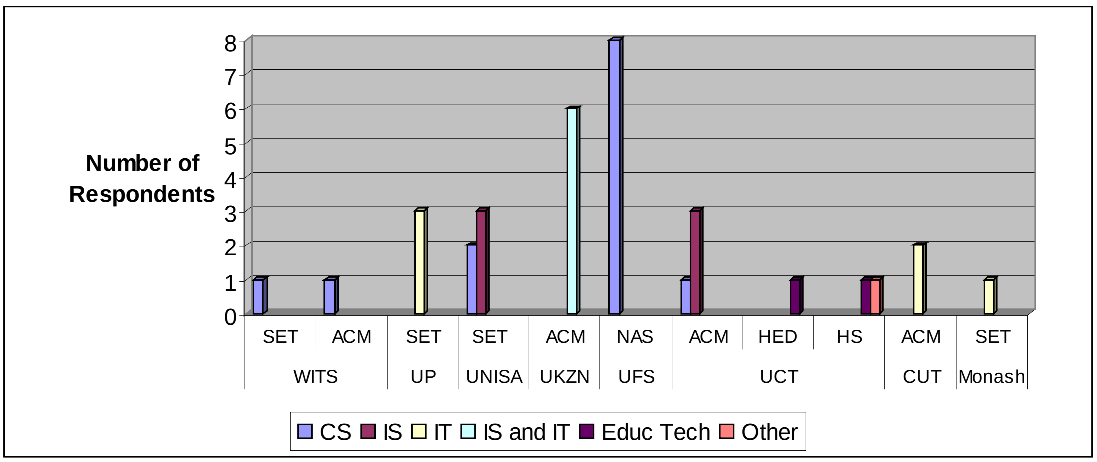
Table 5.2: Distribution of participants by institution, faculty and school/department

		School/Department						
Institution	Faculty	CS	IS	IT	IS and IT	Ed Tech	Other	Total
WITS	SET	1						1
	ACM	1						1
UP	SET			3				3
UNISA	SET	2	3					5
UKZN	ACM				6			6
UFS	NAS	8						8
UCT	ACM	1	3					4
	HED					1		1
	HS					2	2	4
CUT	ACM			2				2
MONASH	SET			1				1
TOTAL		***13***	***6***	***6***	***6***	***3***	***2***	***36***

SET = Science, Engineering and Technology	ACM = Accounting and Management Studies	NAS = Natural Sciences	HED = Higher Education	HS = Health Sciences

Table 5.2 shows that the largest number of respondents came from CS (thirteen), while IS and IT had a total representation of eighteen respondents each. The response level from each university included nine participants from UCT (University of Cape Town), eight from UFS (University of Free State), six from UKZN (University of KwaZulu-Natal), five from UNISA (University of South Africa), three from UP (University of Pretoria), two each from WITS (University of the Witwatersrand) and CUT (Central University of Technology), and one from MONASH University of South Africa.

Figure 5.2 summarises the distribution of respondents in terms of Institution, Faculty and School/Department. Schools were categorised as set out in the graph:



UCT = University of Cape Town	**UFS = University of Free State**	**UKZN = University of KwaZulu Natal**	**UNISA = University of South Africa**	**UP = University of Pretoria**	**WITS = University of Witwatersrand**	**CUT = Central University of Technology**

Figure 5.2: Distribution by institution, faculty and school/department

5.1.4 Participation: by university

Figure 5.2 also indicates that, in total, a sample of 36 South African tertiary academics volunteered to participate in this study. Sixteen (44%) of them were current users of e-assessment systems while 20 (56%) were potential or future users. Of the sixteen respondents who indicated that they do currently make use of these tools, two were from UP, two from UNISA, three from UFS, seven from UCT, one from CUT and one from MONASH University of South Africa. These sixteen current users represent six different teaching units. This is a very small number and indicates low-level usage of e-assessment tools by South African CS, IS and IT academics at the time of the study, late 2009 and early 2010.

5.1.5 Quantitative findings of Study 1: users of e-assessment

This section discusses the findings of Study 1 that lent themselves to quantitative data analysis. The questions are presented together with the results obtained.

The first aspect investigated, asked the sixteen participants who were users of e-assessment (adopters) to indicate what online testing tools they use. Very few of the tools mentioned in Table 5.3 below correspond with those identified in the literature (Section 2.5.5).

Table 5.3: Tools currently adopted

Online Testing Tool	Frequency	Percentage (%)
Sakai	3	19
Vula	4	25
CISCO	1	6
Blackboard	1	6
Moodle	1	6
CompAssess	3	19
Other	5	31

Statistical analysis of these tools crossed with school/departmental classification shows that significantly more than expected (Fisher's exact (N=15) = 7.376, p=.042) participants from an IS classified department/school use Vula, as shown in Table 5.3. Vula was limited to the University of Cape Town, where it is UCT's online collaboration and learning environment, used to support UCT courses as well as other UCT-related groups and communities. Vula was jointly developed with other universities worldwide as part of the Sakai Project. Four adopters at UCT utilise Vula's in-built e-assessment tool.

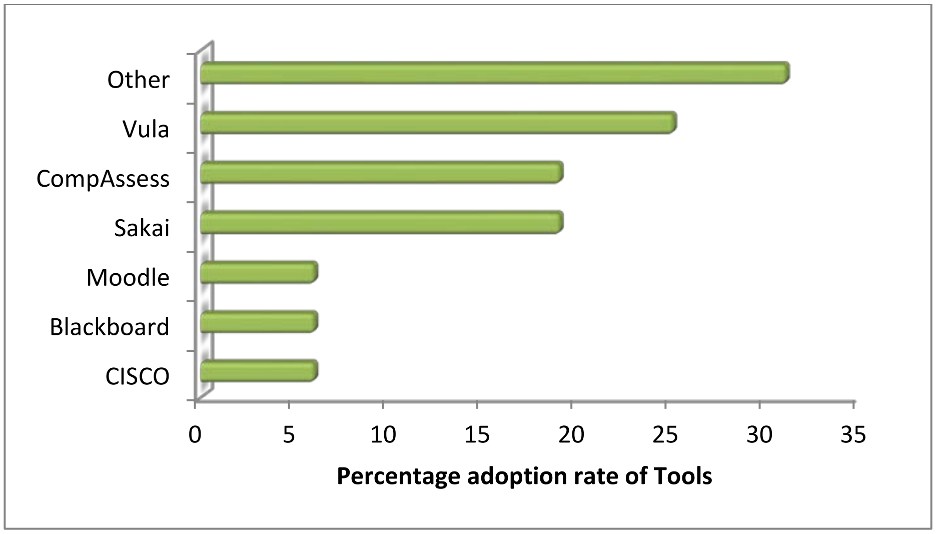


Figure 5.3: Tools currently being used

Figure 5.3 presents the various e-assessment tools and systems currently being used by South African academics in CS, IS, and IT, as reported by the sixteen respondents who are users. Open source tools (Vula and Sakai) are highly adopted. A local-grown e-assessment tool, CompAssess, features higher up on the list, when compared to the e-assessment tools within Blackboard and Moodle LMSs. The tools mentioned under 'Other' include 'Self-Assessment My UNISA'; 'Home-grown automated marking systems'; 'Umfundi and Click UP'; 'Tests, quizzes and examinations' or testing tools that are part of various learning management systems.

The next question enquired how many years users of e-assessment had been adopting online assessment tools. The three categories provided were 1-2 years; 3-5 years; and more than 5 years. Table 5.4 indicates that five of the teaching units surveyed had used e-assessment tools for more than five years; a further five units had done so for three to five years; and six units are new users who had employed these tools during the previous one to two years.

Table 5.4: Number of years for which e-assessment had been adopted

		Years in Use		
		1 - 2 years	3 - 5 years	>5 years
School/ Department	CS	1	1	4
	IS	2	2	0
	IT	1	1	1
	Other	1	0	0
	Ed Tech	1	1	0
Totals		**6**	**6**	**5**

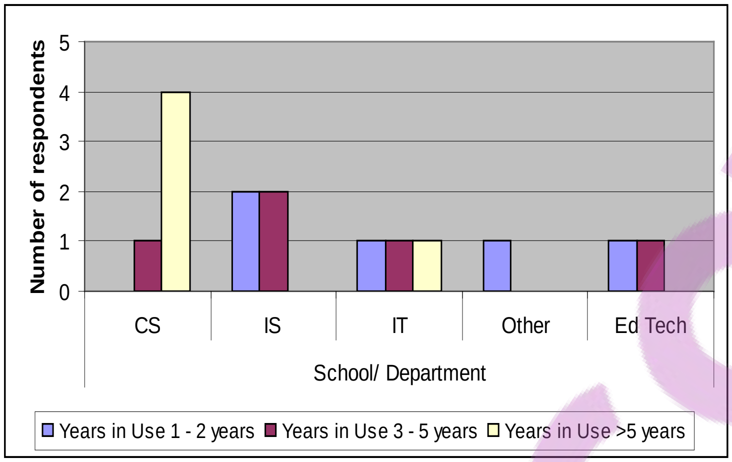


Figure 5.4: Analysis of usage of e-assessment by school/department classification

Although Figure 5.4 shows that more of the CS departments had used the tools for more than five years, this relationship is not significant. IT departments were also among the early adopters of e-assessment. However, with equal distribution in all three time categories, IT departments seem to be kindling interest in e-assessment, following the success of these early adopters.

Which types of questions were supported and used by the sixteen actual users in the tools they adopted, was the next concept explored. The tools being used incorporated questions and items in forms such as basic Multiple choice, True/False, Fill-in-the-blank, Hotspot, Matching, Diagram/Video clips, and Short Answer questions. This covers most of the types outlined in the literature in Section 3.1.5, hence confirming in practice the items presented in the Literature Study. The percentage of usage of these types of questions is summarised in Table 5.5.

Table 5.5: Question types adopted by users

Types of questions	Usage Ranges			
	0 - 10%	11- 50%	51 - 80%	81 - 100%
MCQ	8	2	5	1
T/F	12	4		
Fill in blanks	14	2		
Hotspot	15		1	
Matching	14	2		
Diagram/Video Clips	15	1		
Short Answer	12	4		

As summarised in Table 5.5, eight adopters utilise MCQs for 0 to 10% of their e-assessments; two users for between 11 and 50%; five users for between 51 and 80%; and one user for between 81 and 100% of their e-assessments in a module which they teach. Likewise, Table 5.5 also depicts the usage ranges of the varying question types available in e-assessment tools. Multiple choice and True/False type questions, which are the most basic formats of questions for assessment, are the most supported and used. Significantly more than expected participants used the Short Answer questions, despite these having to be marked manually.

The use of e-assessment for formative and summative assessment was the next concept considered. Significantly more of the adopters than expected (p=.010) responded affirmatively to using e-assessment for formative assessment, as presented in Table 5.6. Eleven of the users reported employing it for formative assessment, and nine for summative. Four of them used it for both formative and summative.

Table 5.6: Summative vs formative usage

Type of Assessment		Frequency	Percentage (%)
Summative (9)	Yes	9	56
	No	5	31
	Missing	2	13
Formative (11)	Yes	11	69
	No	2	13
	Missing	3	19

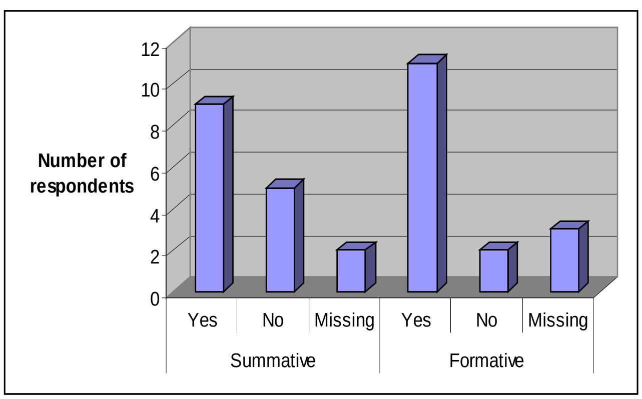


Figure 5.5: Summative vs formative usage

The extensive use of e-assessment for formative assessments, as depicted in Table 5.6 and Figure 5.5, is notable. It shows that academics are encouraging the adoption of e-assessment for practice and revision, in addition to using it for formal testing.

The next question explored the level of study and number of students for which e-assessment was used by the adopters.

Table 5.7: Levels at which e-assessment is being used

Year of study	Frequency	Usage percentage (%)	Average class size	Range of class sizes
First year	**13**	**81**	1492	50-5000
Second year	**5**	**31**	505	120-800
Third year	**4**	**25**	215	50-600
Postgraduate	**3**	**13**	13	10-15

Table 5.7 indicates that the highest usage of e-assessment tools occurred at the entry level, among first-year students, where student numbers were at their highest. Eighty-one percent (81%) of the users employed e-assessment for their first level, in contrast to only 31% at second level and 25% at third level, with low use at postgraduate levels. This aimed at providing an understanding about whether academics are adopting e-assessment just to manage large classes and lower-level students. The data provided valuable insight, indicating that adoption was prevalent beyond the domain of first year and large classes. MCQs had some added value even for postgraduate students.

5.1.6 Qualitative findings of Study 1: users of e-assessment

The second part of the questionnaire contained open-ended, discussion-type questions from which qualitative data emerged. These investigated benefits and disadvantages associated with e-assessments and served to determine users' satisfaction with the use of e-assessment tools. Both users (16: 44%) and non-users (20: 56%) were asked to answer these questions.

Manual analysis was undertaken to extract themes from the textual responses given by participants. Their qualitative responses are presented separately in Section 5.1.6 for users and Section 5.1.7 for non-users. The results corresponded closely with the secondary data from the literature (Sections 2.5.7 and 2.5.8).

The first open-ended question of Study 1 asked participants their opinion on whether e-assessment is more effective than the traditional methods of assessment. Confirmed users stated their belief that e-assessment is more effective than traditional forms of assessment. Eighty-one percent of the users fell into this category. They gave reasons for this belief and mentioned strengths of e-assessment.

The tables of qualitative data that follow were obtained by quantification of qualitative data.

Table 5.8: e-Assessment vs traditional methods of assessment

Theme	Frequency	Percentage (%)
Feedback	12	74
Improved marking	11	69
Management and Control	5	31
Time Savings	2	13
Better administration	2	13

As summarised in Table 5.8, spontaneous textual responses, without any prompts, indicated that:

- feedback is available immediately to students,
- the marking/grading process is faster, more accurate, and always consistent,
- automated testing allows for better management and control of large classes,
- e-assessment saves time, and
- the approach provides easier administration of formative and summative testing and other results.

Three respondents (19%) indicated they were unsure which of the two methods is better, because both methods had their own distinct advantages and disadvantages.

"I am not convinced that online testing tools are any better than traditional tools in terms of genuine academic effectiveness. In fact, there is always the danger that the technology used becomes the subject of the exercise, rather than the material being examined. The advantage that it does bring is the ability to provide standardised tests for large numbers of students without the additional load associated with manual marking. I don't believe that this provides justification in itself however". This statement by Respondent 5 of Study 1, S1(5), indicates that the assessment material must still be the focus, and the technology merely a support mechanism for administrating the assessment, rather than the technology being the focus.

The benefits associated with the adoption of e-assessment to both students and educators respectively, were investigated.

Table 5.9: Benefits of e-assessment outlined by adopters

Benefits to students	Frequency	Percentage (%)	Benefits to educators	Frequency	Percentage (%)
Immediate feedback	12	75	Time savings	13	81
More assessments	8	50	More assessments	10	63
Accessibility	6	38	Better management	9	56
Simulation	3	19	Efficient statistical analysis	8	50
Marking Consistency	2	13	Less administration	5	31
			Easier cheating detection	3	19

Respondents outlined the following direct benefits, as tabulated in Table 5.9:

- For students –
  - availability of immediate feedback,
  - access to more tests, with a wider variety of questions (due in part to question banks synthesised by the educators),
  - convenient and easy access to assessments,
  - the facility to work in a simulated environment, and
  - uniformity and consistency in marking.

- For educators –
  - economy of scale, in that less time has to be spent on marking, allowing more time for other academic activities,
  - more opportunities to assess students,
  - easier management of large classes,
  - better analysis of student performance,
  - reduction in the major administration associated with managing student records, and
  - assistance in minimising cheating through random generation of questions.

Respondent S1(8) provided useful comments on the value of good e-assessment questions by stating that "….. online testing is very good for assessment of the understanding of concepts – especially if the questions can be asked in terms of multiple choice questions that can be assessed automatically. Although there are many options of type of questions in many of the tools, I prefer multiple choice and short answers. When you expect students to write paragraphs online it is also very useful, since you don't have to decipher bad handwriting".

Respondent S1(2) stated that "to enhance student learning, other forms of assessment are (also) needed, as not all students learn in the classroom". This was supported by Respondent S1(16)'s comment that e-assessment "should always be used in conjunction with traditional set papers. They should carry a lower weight".

The positive comments listed above correspond closely with the advantages identified in the literature (Section 2.5.7).

Barriers hindering the effective use of e-assessment were explored in Question 12 of Study 1. These are summarised in Table 5.10.

Table 5.10: Barriers to e-assessment adoption

Barriers	Frequency	Percentage (%)
Time-consuming	6	38
Technical Issues	5	31
Infrastructure Problems	4	25
Level of questioning	2	13
Non-completion of assessments	1	6
Increase in cheating	1	6
Inflexibility	1	6

Barriers identified, as presented in Table 5.10, regarding the adoption of e-assessment tools include:

- it is time-consuming to build a comprehensive set of good questions,
- technical issues versus ethical challenges, for example, student test submissions maybe lost either deliberately or unintentionally,
- infrastructural issues:
  - insufficient availability of computers,

    - technology flaws and bandwidth problems,
    - requirement for a highly competent systems administrator,
- e-assessment is a limiting approach, since most questions are on a low level and do not test insight into learning content.
- some educators are resistant to change, happier to stay with the tried and traditional methods of assessment, despite knowing the limitations,
- students may use the 'failure of technology' as an excuse for not completing an assessment, and
- certain disciplines have their own unique needs, which e-assessment tools do not meet, due to their inflexibility.

In addition, Respondent S1(8) commented that "One serious drawback of multiple choice questions (even in written papers) is that you get students who have mastered the skill of guessing correct answers. Online testing can also not be easily used for assessing the student's ability to be creative or their skill level with regard to the ability to program. I find that if a class is not too big the traditional written papers and practical programming projects are still better ways to assess students".

Expanding on the time issue, Respondent S1(14) stated that "Most lecturers are so pressured by tuition loads and research commitments that they do not have the time to spend on initiating this – even though it could save them time in the long run". This point also draws attention to the need for institutions to put systems and technical support in place, so as to remove this task from the educators.

These barriers outlined by the respondents show a close correspondence with the disadvantages and constraints identified in the secondary data from the literature (Sections 2.5.8 and 2.5.9 respectively).

Participants' opinions on the usefulness of e-assessment tools that are web-based compared to non-web-based tools, was the next aspect studied. Eighty-eight percent (88%) of the respondents indicated that they find web-based e-assessment tools more useful than non-web-based tools. Their reasons are summarised in Table 5.11.

Table 5.11: Web-based e-assessment advantages

Web-based systems	Frequency	Percentage (%)
Question variety	13	93
Equal opportunity	6	43
Minimal resources required	4	29
Less intimidating for some students	4	29
Accessibility	3	22

Participants who indicated their preference for web-based e-assessment tools explained that:

- examiners can present high quality pictures and diagrams by means of colour screenshots,
- they enable a distributed (anytime, anywhere) approach to support additional teaching, communication and assessment,
- they provide better management of large classes, yet utilise less resources (once the system has been acquired),
- students feel more comfortable in an e-environment,
- they support educators in setting exams for different groups of students at different times. Inter-student communication is limited, because the students cannot take question papers away, and
- they give distance-students (distance education is increasingly common) an equal opportunity to access a greater pool of questions on a regular basis.

In addition to the above, Respondent S1(1) clearly felt that both web-based and non-web-based e-assessment tools are good. Supporting this sentiment, Respondent S1(7) said that "if the 'testing' is purely for helping the student in assessment to see if they understand, then web-based would be ideal as the student can do the testing when and where they like. However if it's for formal testing then one cannot have a student answering at some unknown place since one would not be sure how much help the student is receiving".

When asked which of the two types of e-assessment tools, namely web-based and non-web-based, they see themselves adopting in the future, of the 16 respondents, 12 (75%) selected web-based. Their reasons included:

- it depends on the internet speed available at the institution,
- web-based e-assessment tools allow one to set exams for different people at different times (if one trusts them not to communicate). Furthermore, one can use pictures and diagrams very effectively in colour. It also permits formative assessments to be undertaken anytime, anywhere, and
- these tools enable a distributed (anytime and anywhere) approach to support additional teaching, communication and assessment.

Finally, views on the contribution of online testing tools to increasing the ease of use and accessibility to students, were gathered and are summarised in Table 5.12.

Table 5.12: Ease of use and accessibility

Themes	Frequency	Percentage
Longer time frames	8	50
Flexibility	6	43
More and frequent assessments	3	19
Accessibility	2	13
Assignment submission easier	1	6

These views were explained as follows:

- test times can be spread over longer periods for example, a 24-hour window, instead of all at the same time and same place. This facilitates more tests more frequently/regularly,
- online testing provides students with the flexibility to access the assessment in their own time,
- the educator can prepare many quizzes without increasing his/her marking load, to enable students to better prepare for their summative assessments,
- accessibility is an issue as many students do not have adequate off-campus access, especially to the Internet, and

- distance students can complete and submit their assignments online, both MCQ and "full" assignments without sending the assignments to a regional office. The system is immediately updated with the assignment, and allocated marks are available immediately.

A noteworthy general comment made by Respondent S1(15) is that “regarding MCQ type of assessment material, we should build up a database of such questions to share across institutions. In my subject area there is such a database of really excellent MC questions, but it's costly. I have often wondered about building an automated system to take PowerPoint files (for example) and automatically generate simple MCQ and True/False questions from these. Not for assessment purposes (would be too easy) but as a Duly Performed (DP) requirement to ensure students keep reading the material during the course”.

This suggestion clearly calls for greater collaboration amongst academics in the same discipline across institutions, to develop and share questions that can be used in the same subject area, irrespective of the textbook prescribed. This is especially useful where the subject material tests underlying principles.

5.1.7 Qualitative findings of Study 1: non-users of e-assessment software

This section presents a discussion on data obtained from the remaining 20 participants in Study 1, who were non-users, but potential users, of e-assessment. Their responses to the open-ended questions in particular, provided insight into why they had not yet adopted e-assessment.

A variety of reasons for non-usage of e-assessment software were given by the 20 (56%) respondents who are non-users. Some gave more than one reason. These reasons are summarised in Table 5.13.

Table 5.13: Reasons for non-usage

Reasons	Number of Respondents	Percentage (%)
Traditional methods adequate	8	40
Institutional lack of expertise/knowledge	4	20
Personal lack of expertise/knowledge	3	15
Lack of finance in the institution	0	0
Other	10	50

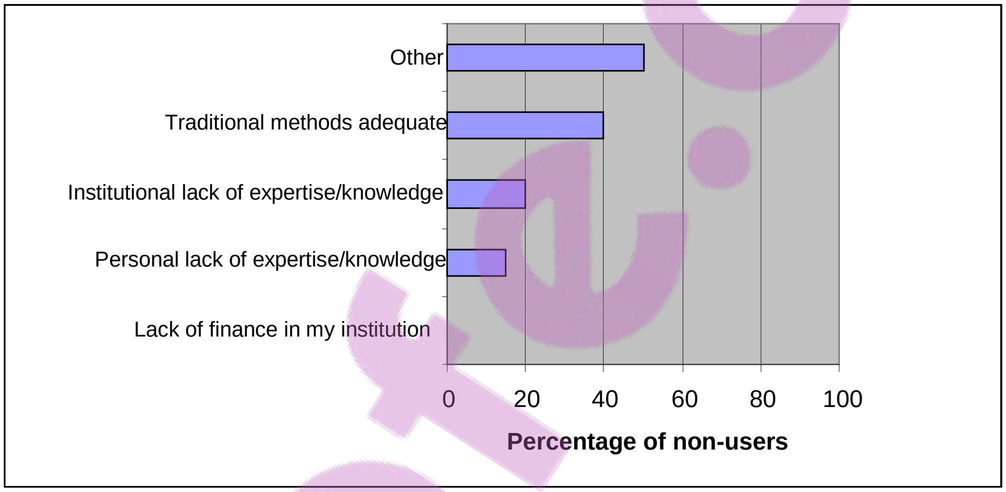


Figure 5.6: Reasons for non-usage

The largest category of reasons for non-usage was 'Other', as illustrated in Figure 5.6, which included issues such as:

- insufficient time provided for academics to learn the software,
- lack of training on the available software,
- not all students have adequate access to the Internet,
- many students encounter bandwidth problems,
- management decisions prevent academics from adopting e-assessment tools, and
- e-assessment systems unavailable at the university or in the department.

As with users of e-assessment in Section 5.1.6, the non-users were asked their opinions on the usefulness of e-assessment tools that are web-based compared to non-web-based tools. Thirteen (65%) of the respondents indicated that they would find web-based e-assessment tools more useful than non-web-based tools. These reasons are summarised in Table 5.14.

Table 5.14: Web-based e-assessment advantages

Web-based systems	Frequency	Percentage (%)
Accessibility	6	43
Familiar and intuitive web interface	4	29
Integration with the Institutional LMS	3	22
Remote testing	2	13
Scalability	1	7

Participants who indicated their preference for web-based e-assessment tools explained that web-based tools:

- are more accessible and easy to use,
- are synonymous with the IS discipline, since they provide the familiarity of an intuitive web-type interface for students when compared to non-web based applications, and therefore have a longer life span,
- can be linked to the institution's LMS,
- could be adopted depending on the nature/level of the course, since not all academic material is suited to MCQ testing, and
- enable remote testing and are scalable as student numbers increase.

Table 5.15: Non-web-based e-assessment advantages

Non-web-based systems	Frequency	Percentage (%)
No web reliance	3	25
Validity of screen designs	2	12
Better security	1	8
Less bandwidth requirements	1	8

A small number of users (seven) indicated their preference for non-web-based tools. Some of the reasons in Table 5.15 include:

- less reliance on web access,
- the validity of screen designs can be tested with inbuilt tools in the e-assessment software,
- better security, and
- less bandwidth requirements.

A further comment made by Respondent S1(8) was that both web-based and non-web-based tools could be used depending on the nature of the module. If the students are required to motivate their answers by using theory from the text and other sources, together with the case study, a combination of web-based and non-web-based tools could be adopted.

For future adoption, web-based tools seemed to be favoured, with 65% of the non-user participants expressing interest in adopting them in the future. Some of the reasons were:

- these tools make it easier for the student to complete the task in his/her own home. Furthermore, it is easier for the educator to distribute the assessment. Also, most students would be comfortable with such tools, due to the familiarity of the interface, and
- they are easier and more convenient to implement.

A further comment made was that "in courses that are all very practical a testing tool that just asks multiple choice or short questions, is of little use" S1(8). He further added "however, a tool, whether web-based or non-web-based, that can be used to assess the practical concepts of the courses, would prove invaluable".

5.1.8 Conclusion and summary of Study 1

Prior to this study, the researcher had limited knowledge of the extent of usage of e-assessment tools at tertiary institutions in South Africa, as well as little understanding of the nature of this use with regard to the types of questions and items tested, the level at which such tests were used, and the like. Hence, this study, which was the first in an envisaged set of three (see Figure 4.8 and Figure 5.1), aimed to establish a broad context as the foundation for subsequent research. The researcher realises that this 2009/2010 study does not provide a comprehensive or complete quantification of the use of e-assessment in South African tertiary institutions, but it provides a clear indication of the nature and context of the usage.

According to the survey responses, of the nine institutions represented, usage appeared to be concentrated in six tertiary institutions: UCT, UFS, UNISA, UP, CUT and Monash, with more users being Computer Science academics, who tended to adopt these tools earlier than Information Systems and Information Technology users. In some CS academic units, the tools and systems had been deployed for more than five years. Although the actual extent of usage is low, it is steadily on the increase, as evidenced in Table 5.4.

To summarise the findings, there were 36 respondents from nine institutions, of whom sixteen were regular users of e-assessment and testing, while 20 were potential users. The systems were used more for formative than for summative assessment. Most usage occurred in cases of high student numbers, that is, first-level classes with numbers ranging from 50 to 5000. Questions most frequently used in assessment were multiple choice questions, and true-or-false questions. This indicates a need for academics also to ask questions that require use of higher order thinking skills (HOTS) and the upper levels of Bloom's taxonomy, as discussed in Section 3.1.3 of Chapter 3.

Deployment of tools is either limited, or supported by, institution-wide policies. Although a number of academics make ad hoc use of e-assessment and Computer-Based-Training (CBT), certain institutions have official policies and procedures, and promote established practices. In such cases, there are dedicated laboratories for computer-based testing and administrators to manage testing sessions. The results of summative assessment in the form of tests and exams are recorded automatically on students' academic records, as well as on class records.

Despite the small number of users, a variety of tools was adopted, including Sakai, Vula, CISCO, Blackboard, Moodle, and CompAssess. Assessment was also conducted using the Self-Assessment tool on the myUnisa learning management system (LMS); Umfundi and Click UP, which are custom-developed automated marking systems; and tests/quizzes on various LMSs.

To determine satisfaction, the questionnaire probed the sixteen established users on their perceptions of benefits and disadvantages. The qualitative open-ended responses (Section 5.1.6) are mainly in line with the secondary data from the literature study (Chapter 2). Though few in number, these established users are, in the main, convinced users. Eighty-one percent (81%) of them believe that e-assessment is more effective than traditional forms, and motivated this by giving their reasons. They expressed satisfaction with the concept of e-assessment, and pointed out advantages for both educators and students. They also addressed disadvantages and barriers resulting from the use of such tools.

Despite her belief that e-assessment of MCQ-related questions has many strengths and benefits, the researcher acknowledges the drawbacks and concurs with the reservations. In particular, she takes cognisance of the issue identified in the literature and raised again in the survey, that questions from the MCQ family, whether online or paper-based, are not appropriate to test all forms of knowledge and learning. Moreover, it is very difficult to design questions that assess insight and higher order thinking skills. This was addressed in Section 3.1.3.

This study has served to establish the nature and extent of usage of electronic assessment tools in CS, IS, and IT academic units at South African higher education institutions, and investigated the satisfaction of the users of such tools. It set the scene for further studies in this action research series.

5.2 Study 2

This was an interview-based study which presents both quantitative (Section 5.2.3) and qualitative (Section 5.2.4) findings. Participants in this study were all users of e-assessment.

5.2.1 Introduction to Study 2

Study 2, which was conducted during late 2010 and early 2011, involved interview follow-ups to Study 1, targeting users of e-assessment. Like Study 1, it aimed to answer Research Question 1. The interviews addressed the same issues as Study 1, but more in-depth. Through these interviews, some additional criteria were gleaned for inclusion in the SEAT Framework, as presented in Table 5.35.

The interviews included a large quantitative component, which is discussed in Section 5.2.3, with the purely qualitative concepts discussed in Section 5.2.4. As in Study 1, manual analysis and coding were undertaken to extract themes and patterns from the textual interview transcripts. Although there were 72 interviews, the responses lent themselves to manual analysis.

Since poor responses were received in Study 1 to e-mail requests for participation, telephonic and personal interviews were used in Study 2 to gather many responses in a shorter period. A summary of Study 2 is presented in Table 5.16.

Table 5.16: Summary of Study 2 as outlined in Figure 5.1

Study 2 (May 2010 – February 2011) – Actual e-assessment Usage	
Respondents	72 respondents (68 included in the quantitative analysis) from both Computing and Non-Computing-related departments in 11 universities
Data Collection	Personal/Telephonic Interviews – Interview questions (See Appendix D) Informal observation, where possible
Data Analysis	Quantitative – basic statistical analysis Qualitative – content analysis
Purpose	To further understand actual usage of e-assessment at South African academic institutions and to identify respondents' opinions on criteria for evaluating e-assessment systems.

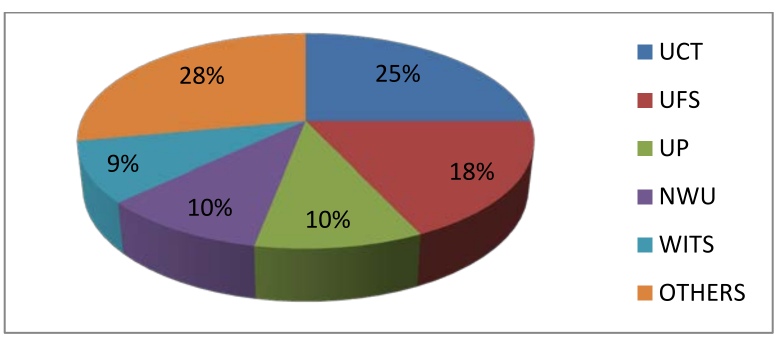
5.2.2 Interview participants

Seventy-two (72) interviews – some personal and most telephonic – were conducted with South African academics. The informal observations where participants volunteered to demonstrate the e-assessment system they adopted, were an enhancement to the interviews, and did not serve as a formal data collection method. Eighteen (18) participants who had participated in Study 1 were also participants in Study 2. Just under half of the interviewees were Computing participants. Participants in Study 2 were recruited by:

- reusing the participants of Study 1,
- the researcher acquiring departmental staff lists and personally making contact with academics who had not participated in Study 1 and inviting them to contribute, and
- referrals from participants in Study 1.

The sampling methods were thus a combination of convenience sampling, purposive sampling, and snowball sampling (Cohen et al., 2011).

The researcher was also referred to four local e-assessment developers, who were interviewed for information on the design requirements of South African academics. Their contributions were particularly valuable in relation to benefits and problems of e-assessment software, as well as the mention of additional functionalities and features used or requested which contributed to the identification of evaluation criteria for Table 5.35.



UCT = University of Cape Town	UFS = University of Free State	UP = University of Pretoria	WITS = University of Witwatersrand	NWU = North-West University

Figure 5.7: Institution participation

Four responses were only partially included, that is, they are not incorporated in the quantitative reporting, but their open-ended responses are reported in the qualitative analysis of Study 2 (Section 5.2.4), as they were non-academic. These respondents included support staff, an administrator, and developers. Therefore, as indicated in Table 5.16, 68 respondents were included in the quantitative analysis.

The 68 participants included in the quantitative analysis of Study 2 (Section 5.2.3) represented eleven institutions (including two universities of technology) from eight provinces. As illustrated in Figure 5.7, 25% were from UCT and 18% from UFS [Qwa Qwa and South Campuses]; UP (10%); NWU (10%) [Vaal and Potchefstroom campuses]; WITS (9%); and 28% were smaller numbers coming from the six other institutions.

The interview protocols are presented in Appendices D1 and D3. The questions in both the personal and telephonic interviews were not always asked in the same order, as the interviews were unstructured, to allow the participants to voice their opinions freely. The personal interviews also gave the researcher the opportunity to observe the actual use of the e-assessment tool being described by the participant. Furthermore the questions in the personal interview schedule (Appendix D1) differ slightly from those in the telephonic interview (Appendix D3). The questions in the personal interview schedule consisted of the basic questions regarding participants' usage of e-assessment tools, while the telephonic interview questions also included questions from Study 1, in addition to the questions focusing on e-assessment tools usage. As a result, the presentation of data in Sections 5.2.3 (quantitative findings) and 5.2.4 (qualitative findings) is structured according to logical categories rather than following the sequence of questions in the interview schedules.

5.2.3 Quantitative findings of Study 2

The interview study supplemented and extended the questionnaire adopted in Study 1 by providing further insights into the adoption patterns of e-assessment, with data from a larger group of participants, namely 68 instead of 36. The data obtained in certain interview questions is quantified, but many of the quantitative findings are supported by qualitative comments by participants.

As in Study 1, the first set of questions set the context of e-assessment usage. According to Costagliola & Fuccella (2009), most online testing modules are components of general purpose learning management systems (LMSs). Hence, the researcher explored whether the tool adopted was part of the Institution's LMS and found this to be the case in South Africa as well, with 74% of the interviewees using the tool built into their university's LMS. Only 26% of academics adopted 'pure' e-assessment tools. These tools included standalone e-assessment tools like Umfundi, Top Class, CompAssess, SAM and web-based e-assessment tools like Clickers, Respondus and HotPotatoes.

The findings are presented in Table 5.17.

Table 5.17: Tool embedded in the Learning Management System (LMS)

Tool type	Frequency	Percentage (%)
Embedded in LMS	50	74
e-Assessment	18	26
TOTALS	*68*	*100*

Respondent S2(8) stated that in his discipline, online assessment is embedded in Blackboard LMS but "although most online assessment is undertaken in postgraduate modules, there is increasing usage of these tools for undergraduate programs …… requires less bandwidth since it is part of the LMS". Respondent S2(48) indicated that most tools he has adopted are "within the LMS. In 2010 [he used] BlackBoard Vista8 and also Moodle 1.9, within the LMS – with features such as the assignment submission tool, discussion forums, quiz tool, Turnitin (plagiarism) and external blogs being adopted".

Next, participants were requested to indicate what e-assessment tool(s) they use(d) and for how many years they had adopted the tool. As outlined in Table 5.18, the e-assessment tools adopted by these participants included Umfundi/Top Class, CompAssess, HotPotatoes, SAM, Clickers and Respondus. These were different to those named in Study 1, thus providing a larger variety of e-assessment tools in use in South Africa.

Table 5.18: Tools adopted

Tool	Frequency	Percentage (%)
Umfundi/Top Class	8	44
CompAssess	4	22
HotPotatoes	2	11
SAM	2	11
Clickers	1	6
Respondus	1	6

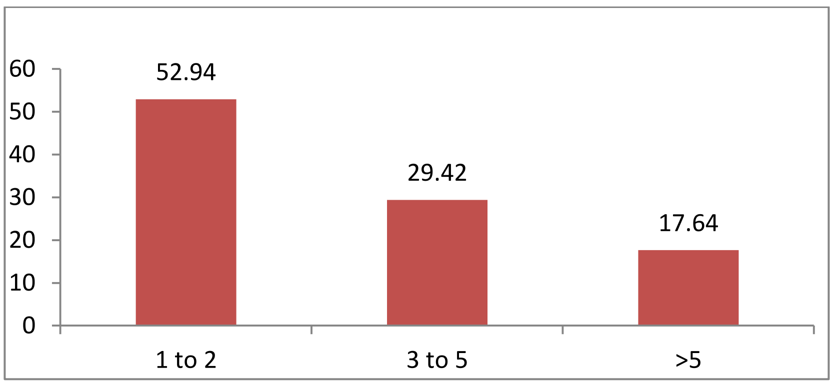


Figure 5.8: Number of years e-assessment adopted

Figure 5.8 shows that of the 68 participants, 52.94% were recent adopters of e-assessment, having had one to two years experience with it. Only 17.64% had used it for more than five years. This indicates that more than 50% were new users of e-assessment, demonstrating a trend of increased usage.

The next set of questions investigated whether participants used a single tool consistently between 2009 and 2010, or whether they used different tools for different purposes, or whether they changed their preference, and for what reason.

Table 5.19: Change in tool adoption

Change indicated	Frequency	Percentage (%)
Yes	10	15
No	58	85
TOTALS	*68*	*100*

It emerged, as depicted in Table 5.19, that 85% (58) of participants had used the same tool for the previous three years. Of the 15% (10) who changed tools, five did so due to institutional changes, while the other five adopted a different tool due to personal dissatisfaction and interest in exploring other options and question types not supported by the tool.

At one institution, it was explained that "when we started we were one of the first, but the university had made it clear that this was their chosen product and that they would persuade as many departments as they could to use it. Therefore this online assessment tool (selected by that university) was adopted (by this academic)" Respondent S2(62).

Another, Respondent S2(36), stated that "our University always used a proprietary LMS that was developed in-house, but the last couple of years we've been using SAKAI which is of course open source, so we were forced to adopt the online assessment tool within SAKAI".

Further understanding of the usage patterns of e-assessment tools required the researcher to investigate what Institutional and School/Department policies were existent at each of the Institutions represented by the participants.

Table 5.20: Policies on the adoption of e-assessment

Institutional Policy	Frequency	Percentage (%)	School/Department Policy	Frequency	Percentage (%)
Yes	8	12	Yes	0	0
No	60	88	No	68	100
TOTALS	*68*	*100*	*TOTALS*	*68*	*100*

As presented in Table 5.20, 12% of the participants indicated that their institutions have a fixed, university-wide policy on adoption of e-assessment. Academics must conform to the policy and design assessments around it. The other 88% were free to use e-assessment as they wished. None of the interviewees was restricted by any School/Department-wide policy. The adoption or nonadoption of e-assessment within the School/Department is left to the lecturer to decide, but if the lecturer chooses to adopt a particular tool, the decision is supported.

Support for these academics using e-assessment tools is most frequently provided by the University's e-learning unit, as illustrated in Figure 5.9.

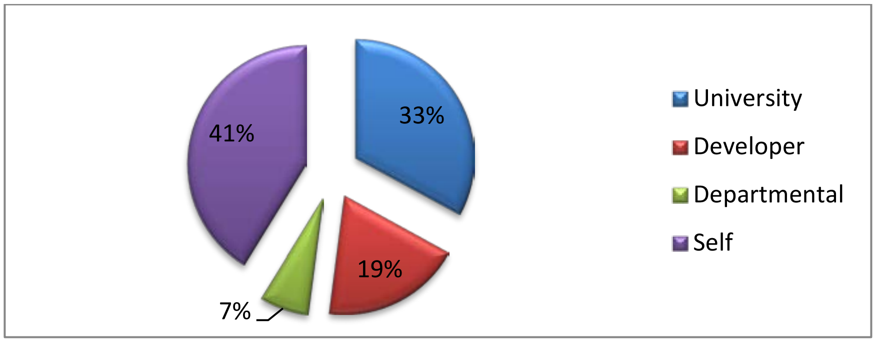


Figure 5.9: Support provided for e-assessment

All the interviewees (100%) indicated they had the full support of management in their department/school for adopting e-assessment.

Interviewees were then asked how much e-assessments contributed to the final mark of a student.

Table 5.21: Contribution of e-assessment to final mark

Contribution	Frequency	Percentage (%)
< 10%	30	44
11-40	12	18
41-80	11	16
81-100	10	15
TOTALS	*63*	*93*

As presented In Table 5.21, for 44% of academics, the e-assessment component contributed less than 10% to students' final marks. This corresponded with their responses on whether the adoption of e-assessment was mainly for formative or summative purposes. Five respondents did not answer this question. Figure 5.10 shows that e-assessment was used by 70.6% of the academics for formative assessment (which did not contribute to the final mark), while 29.4% used it for summative purposes and 28.6% for both. Health sciences and Medical faculties were the largest users of e-assessment for summative purposes.

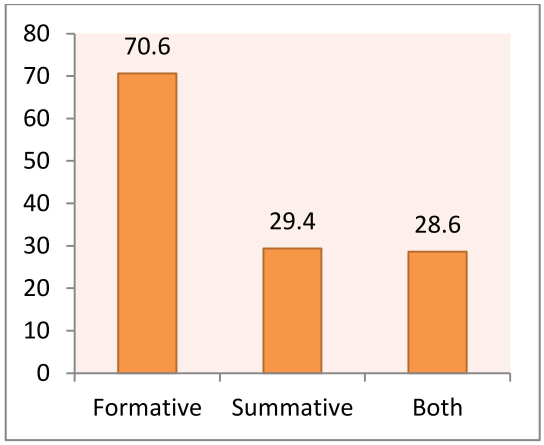


Figure 5.10: Adoption of e-assessment by type

A respondent who adopted e-assessment for summative purposes indicated that although she was keen on also using e-assessment for formative purposes, this is "problematic as the database of questions I have is not large enough, and thus unable to provide too much feedback to students" S2(28).

Quotations from participants who were enthusiastic users of formative e-assessment, now follow:
An adopter of e-assessment for formative purposes indicated that she used e-assessment for "formative assessments (with MCQs) only, on both Moodle and WebCT as I am able to provide comprehensive feedback to students; as well as obtain an understanding of their prior knowledge and current level of learning" S2(24). Another user of formative assessment stated that "these frequent assessments in the form of quizzes, which the student has to complete within a week, are for them to gain more knowledge in the area being tested, as well as force them to interact with the material" S2(10). Respondent S2(33) supported this concept indicating that she uses "online assessment to encourage students to read the chapter before coming in to lectures, to get them to engage with the material". Further, by providing students with the opportunity to "attempt each assessment as many times as they want to, they can improve their learning based on the feedback they received on their attempt" S2(12). "The introduction of quizzes for self-assessment, which

provide the students with immediate feedback without creating a grade is an essential component of their self-learning", said respondent S2(38).

Respondent S2(8), who adopts e-assessment for both summative and formative purposes, indicated that "summative assessments are usually closed book and invigilated, whereas formative assessments are open book and can be taken as many times as the student wants to. The aim is for students to identify gaps in their knowledge. Marks are awarded for formative assessments, but they do not count to their final year mark, but the lecturer can view the marks of each attempt to monitor student progress".

The next aspect studied was the timing of the release of marks from e-assessments.

Table 5.22: Mark release

Timing	Frequency	Percentage
Immediate	59	87
Delayed	9	13
TOTALS	68	100

Marks for these assessments, as tabulated in Table 5.22, are released immediately by 87% of academics via the assessment system, while 13% delay release for checking and moderation, and to prevent copying where questions are repeated in different versions of the same e-assessment taken at different sittings. As S2(54) indicated, "Marks are hidden even though captured on the system, until finalised by moderator" and then they are released.

Interviewees were asked at which levels e-assessments were used, and for what class sizes. Table 5.23 summarises adoption at the various levels. The highest adoption of electronic questions from the MCQ genre occurred at first and second levels (National Qualifications Framework (NQF) levels 5 and 6), with large numbers of students. More than two thirds of the participants used e-assessment at first level.

Eight academics also used e-assessment for postgraduate students, with relatively small numbers (5–25 in a class). The final column indicates the percentage of participants who used e-assessment at those levels. The percentages total more than 100% because some used e-assessment at more than one level.

Table 5.23: Levels where e-assessment is adopted

Level of Study	Number of students in cohorts	Adoption rate - Frequency (number of interviewees who use at that level)	Percentage (%)
First Year	>250	44	67%
Second Year	200-500	12	19%
Third Year	40-120	10	15%
Postgraduates	5-25	8	12%

As illustrated in Table 5.23, "... usage is wide where undergraduate courses have large numbers of students, which is a motivation for adoption due to the marking time being saved" said respondent S2(44). This is supported by respondent S2(18)'s statement that "... online assessment was adopted because of the large number of students and the demanding marking load associated with this". However, some participants (12%) adopt e-assessment for postgraduate students, as shown in Table 5.23. Respondent S2(12) stated that she created "higher-level of MCQs where students have to motivate why they have selected an option, for application in postgraduate modules".

Although e-assessment is commonly adopted for large classes, as indicated in Table 5.23, participants also use this type of testing for 'smaller classes' ranging from 5 to 40 students per cohort. Respondent S2(62) indicated that although he had "approximately 20 students in the class – despite the smaller number of students, online assessment is effective. The turnaround time is quick, painless – no effort from anyone to sit and mark – generally most academics don't like marking. It is also very objective, all subjectivity removed".

The next question investigated how many online assessments were adopted for each module. These results are summarised in Table 5.24.

Table 5.24: Number of e-assessments per module

Number of e-assessments	Frequency	Percentage (%)
1-2	42	62
3-5	18	26
>5	8	12
TOTALS	*68*	*100*

Figure 5.11 illustrates that only 12% use e-assessment for five or more assessments in a module, while 29% use it for three to five assessments, some of them formative. The greatest percentage, 59%, used it for only one or two assessments per module, indicating a balanced distribution of conventional assessment and e-assessment. In 45.1% of cases investigated, the contribution of e-assessments to the final mark was 11– 40%, while only in 11.7% of cases, did e-assessment contribute over 80% of the final mark. This shows a sound balance between conventional assessment methods and e-assessment.

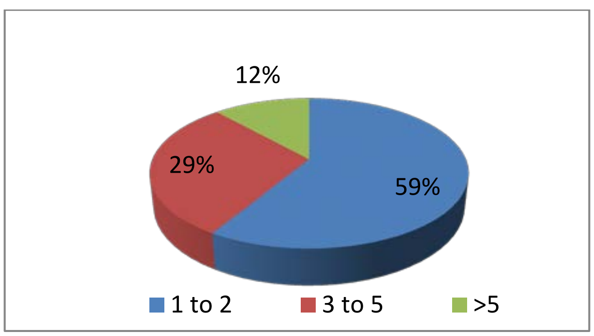


Figure 5.11: Number of e-assessments per module

When asked whether participants who used e-assessment tools required any formal training prior to its adoption, it emerged that 51% of academics did not require any formal training, but were self-taught.

The other 49% attended some form of training or orientation, as summarised in Table 5.25.

Table 5.25: Formal training on e-assessment

Training	Frequency	Percentage
Yes	33	49
No	35	51
TOTALS	*68*	*100*

The administration of electronic assessments was the final quantitative aspect investigated.

Table 5.26: Administration of e-assessment

Administration	Frequency	Percentage
Examination conditions	36	53
Self-administered	32	47
TOTALS	*68*	*100*

With regard to administering assessments, 47% of the academic interviewees did it themselves as depicted in Table 5.26, while 53% had assistance from administrators or invigilators. Forty three percent (43%), not shown in the table, administered their e-assessments under examination-style conditions, in their own laboratories, with invigilators employed to monitor the assessment. The other 57% used e-assessments for students to undertake self-assessment for practice and revision.

5.2.4 Qualitative findings of Study 2

The interviews gleaned qualitative data about interviewees' perceptions of the benefits and drawbacks associated with the adoption and use of e-assessment. The researcher studied the interview transcripts to identify themes and patterns, some of which corresponded with findings from the literature, while others were novel. Furthermore, during the interviews some participants used the open ended questions to discuss what features they liked, disliked, and wanted in e-assessment systems. From these themes and patterns, criteria that were deemed important for inclusion in an e-assessment tool, were also identified. Those criteria were identified as additional criteria for the framework, that is, they did not correspond with those already identified in the literature. They were included in the SEAT Framework, as presented in Table 5.35. As in Study 1, manual coding was undertaken.

5.2.4.1 Benefits of e-assessment

The benefits outlined by the participants in this study fall into ten categories and are supported by quotations, whose authors have been coded so that they remain anonymous – that is to say, their details have been anonymized, as this is the practice in this research.

The benefits are listed in Table 5.27 and discussed below.

Table 5.27: Benefits of e-assessment

	Benefits of e-assessment
1	Improved Feedback to students
2	Improved Feedback to educators
3	More Consistency
4	Reduced Uncertainty
5	Increased Assessment Frequency
6	Question Reuse
7	Improved Accessibility
8	Better Student Engagement
9	Saves the Environment
10	Higher Levels of Testing

Improved feedback to students – Fast turnaround time is a prime advantage. In formative assessments, detailed constructive feedback can be provided. The correct answer is given, possibly supplemented by information such as page references, hyperlinks to relevant resources, and diagrams. Some interviewees offered suggestions regarding this enrichment feedback. It should be accessible to the student both in the assessment venue and in a portable form, such as a printout or download to a Universal Serial Bus (USB). Moreover, the feedback should be equitably provided to all test-takers, including those who got answers right. S2(44) stated that "assessments with immediate feedback on their performance are very beneficial to students". It provides them with "….. proper learning paths" S(2)35. When constructive feedback is provided, it was found that students "…… engage in the discussion after the assessment, about the answers", enhancing student engagement S2(36).

Improved feedback to educators – e-Assessment tools provide almost immediate feedback about "…. what understanding is taking place and how students are coping with material in the class" S2(28). Further, if questions are grouped according to Biggs (2003) "……..the feedback from the system automatically helps you determine the level of learning of students" S2(24). Moreover, e-assessment can assist the academic "to recognise prior learning" of a student and serves as an "early warning system for first years" S2(32).

More consistency – In automated assessment, there is no subjectivity, bias or impact of a human assessor's emotional or physical state. Nor do different markers assess the same test. In contrast, the responses are judged impartially and objectively by a computer program. As S2(38) indicated, "online assessment helps to achieve consistency in the marking process". This was also indicated by S2(24) who stated that "the main motivation for the use of online assessment was the large classes, so that they could be assessed more fairly". Furthermore, "it is also very objective, all subjectivity removed" S2(14), especially "if you do not understand the student's writing" S2(48) .

Reduced uncertainty – Students are not left to wonder for days or weeks about their mark, because results are available rapidly. A quotation from Respondent S2(34) relates to Points 2 (improved feedback to educators) and 3 (more consistency): "To get proper turnaround and mark papers/assignments quickly and efficiently and return them to students is very difficult. Also, it is often not fair, because we used to hire a variety of tutors to mark. Online assessment helps to achieve consistency".

Related to this is the fact that students are exposed to "… just one system, and assessment style …. therefore they interact with the system well without any training, which is now very intuitive for them" S2(46) and therefore are better prepared for their summative assessments.

Increased assessment frequency – e-assessment allows more frequent assessments without additional marking. However, there is a great deal of work and pilot testing when the initial question bank is created. This work continues over successive years, as questions are improved and added. Where enriching feedback is created and included, there is more time-consuming work for the academic. S2(28) states that "The main motivation for using it, is so I can do continuous assessment on a very regular basis with large numbers of students, without the additional burden of marking". Similarly S2(16) indicated that "online assessment assists us to test these large numbers of students more frequently without the marking burden". S2(18) also stated that "academics use online assessment to reduce the demanding marking time associated with large classes" and "too few staff to cope" S2(8). As S2(10) indicated, "online assessment is effective – the turnaround time is quick, painless – no effort from anyone to sit and mark, generally most academics don't like marking".

Question sharing and reuse – Question banks can be created and questions can be reused over time. Where more than one higher-education institution offers a course with similar subject matter, questions can be shared. S2(8) supported this by stating "… not reinventing the wheel – sharing of questions and collaboration with peers; question banks ...". Most academics who adopt e-assessment have "... a huge test bank which has been built up over many years" S2(36). Thus "the computer will randomly generate the questions, so each student gets a unique set of questions, from the same categories/type" S2(18). By "randomising questions and detractors, academics can minimise student cheating and try to prevent questions from leaking out to other students" S2(6). If the questions are "exhaustively checked prior to saving them to the database, this assists with creating a good database" S2(11). Thus, they can be "reused annually and checking not required extensively each year round" S2(62). In the future, academics could "work towards getting the different faculties and disciplines to work together and collaboratively create an online MCQ bank, and possibly develop a sharing agreement between the various universities" S2(66).

Improved accessibility – There can be time-independence and location-independence. Assessments and practice can run on the Web with 24/7 flexibility and in a location of the student's choice. This is termed 'anytime-anywhere-access'. S2(26) stated that "students can attempt the quizzes anywhere because they are linked to the LMS which is accessible through the Internet". S2(14) indicated that "class tests can be taken from anywhere". They can be accessed from "anywhere in the world; for example, a student on holiday in Italy took her material with her and completed the quiz from there" S2(16). These assessments become more accessible if they are "linked to the LMS which is accessible through the Internet" S2(24). Where assessment is true formative assessment that does not contribute to marks, it can be done by the student in any location – a computer laboratory or kiosk, his/her home, or at a workplace. Where it does contribute, caution must be exercised in allowing test-takers to do assessments anytime-anywhere, because there is no assurance it is their own unaided work. One way of countering this is by timed pauses. Questions remain on the screen for a fixed time, which is insufficient to search for the answer in a book. However, it does not prevent the situation where the student has someone on hand to help answer questions. Assessment that contributes towards the mark should be done in official, monitored venues, with students doing it simultaneously or in sessions to manage large numbers of students.

Related to student accessibility is the fact that "…… external moderation can happen irrespective of where in the world the external examiner is, as long as he is added on as a user" S2(66).

Better student engagement – This overlaps with accessibility, because it relates to students answering questions for practice or to improve their learning. One interviewee mentioned "challenging questions", which stimulate students to engage upfront with the upcoming course material. As S2(64) indicated, "online assessment is used to encourage students to read the chapter before coming in to lectures, to get them to engage with the material". Thus frequent assessments in "the form of quizzes……..forces them to keep up with the material" S2(42). It also "allows them to gain more knowledge in the area being tested". Many students are weak with MCQs, so the formative e-assessments adopted "……… gives them good exposure to MCQs, prior to their summative assessments" S2(32). Also, "online assessment provides students exposure to interaction with technology, instead of just the typical traditional paper-based environments" S2(8).

Saves the environment – "This is a 'green system' – no paper handouts or paper hand-ins, all electronic. They don't write anything. Thus we are saving the environment" S2(12).

Higher levels of testing – e-Assessment "can have higher level of MCQs, from Biggs (2003) hierarchy (Application, Understanding, Comprehension, Justify analysis, Recommend an action), where students have to motivate why they selected an option intervention" S2(44). These questions may "include a full case/story and out of that they are required to draw answers from MCQs. This helps to maintain the standard of the questions required, and at the same time maintain the practicality of adopting automated marking" S2(28). Therefore some adopters of e-assessment "realise that it takes a lot of effort and time to develop good quality MCQs, especially if you want to test higher levels of thinking" S2(18). S2(10) indicated that he "asks the same questions that we would in Short Answer/explanations, but now in an MCQ format to facilitate online marking" S2(64). Extended Matching Items (EMIs) take more effort to create but are much more efficient than MCQs because they allow testing of a broader knowledge area with fewer questions. "None of the items in the list can be discounted as they are all part of the domain" S2(44). However, S2(38) advises that "EMIs must be used in conjunction with A-type questions. They are very useful over the long term since you don't need to keep finding new distracters all the time. All the options are plausible; hence there is no wrong answer. Each new question that you write uses the same list of options. So even though it takes more time to construct it at the beginning, once you've got your list of options which relate to the same aspect, it is very easy to write new questions".

5.2.4.2 Disadvantages of e-assessment

The constraints associated with e-assessment as outlined by the participants in this study fall into eight categories and are supported by anonymised quotations. These disadvantages are listed in Table 5.28 and discussed below.

Table 5.28: Disadvantages of e-assessment

	Disadvantages of e-assessment
1	Infrastructure
2	Literacy Skills
3	Not always suitable
4	Technical limitations
5	Training
6	Resistance to Change
7	Security issues
8	Test Bank

Infrastructure – Despite their keen interest in e-assessment, academics seem to face a similar problem of inadequate infrastructure at different academic institutions in South Africa. S2(5) found that "when many students write the test at the same time, especially when they login simultaneously (examinations with 400-500 students), the test was problematic ……. when the same tests are also run across the campuses simultaneously, logging in all these users simultaneously gives a problem ……… technical infrastructure that was problematic”. Despite some academics adopting e-assessment for formative assessments “formal tests and exams are not done online because we have too many students and not enough computers” said S2(56). Another respondent S2(38) “stopped usage of e-assessment due to limited bandwidth, which resulted in access to the quiz being slow”. He further stated that “the network infrastructure at his institution does not permit us to adopt an Internet-based online assessment tool”. As S2(48) stated, “if we had more bandwidth we could implement e-assessment better”. Old and outdated computers used in some labs are unsuitable for running e-assessment. This was reinforced by S2(32)’s statement that “the equipment students use is quite outdated”. “There is limited space to run the tests, not enough computer labs, therefore we have to run multiple sessions per test. This becomes quite intensive with regard to both human logistics and resources, as it takes about four days to complete an assessment with approximately 1200 students”, said S2(22).

Literacy skills – “Challenges are faced by students who have little or no computer experience” stated S2(64). This is less the case with Computing students, although it can be an issue during early stages. Making Computer Literacy a compulsory module across all faculties in the initial years, can assist students in using online assessment better. When a new tool is adopted, it is first introduced to them in a safe non-test environment, to help them familiarise themselves with it. Hence, if the tools are intuitive, they are more accepted by the students. In the same light, the MCQs in e-assessment can be difficult for students to pass. Sometimes, “questions are designed to be tricky, with four options all looking very similar, difficult for students to identify the correct answer”, suggested S2(14). She continued, “those that study hard will cope, but others not, as all answers would seem plausible”. Students who are not confident with a system “often request you to mark it manually again if they receive a low mark from the online assessment system – but once they see the mark is correct, they fully trust the system”.

Not always suitable – Some courses cannot be assessed using MCQs and “there is a danger that some lecturers will water down their questions to make them suitable for an online assessment tool”, indicated S2(32). Furthermore, practical questions in these systems can be limiting as “some questions allow you to answer a question using any of the methods available, for example, in Word (keyboard shortcut, menu, right click …). However other questions require a particular method to be used, but don't always specify so in the question" stated S2(66). Students are nevertheless forced to answer it in the method associated with the question. "If students try to perform a method which is not allowed by the question, the screen freezes, without even an error message, so the invigilator has to select the question again from the list and allow the student to redo the question", continued S2(66). S2(53) suggested that "besides the management, infrastructure and logistical issues preventing an automated exam, there is a need to have a written component to the module". Thus e-assessments should be used in conjunction with written assessments.

Technical limitations – Participants indicated that, in general, most systems are reliable, yet technical problems occur. "Most problems stem from the students; for example, a student choosing to write an examination on a computer that they have not used before and it malfunctions, or is not set up correctly, or internet access fails, or power cuts occur, or ISP issues during the examination" explained S2(27). Other issues included "embedded media freezing or the freezing of quizzes" S2(34). Assistance and support should be available whatever the problem, be it technical, hardware or network-related.

Training – Most lecturers are not competent to handle the administration of the tool, especially when they have not attended any training. S2(42) referred to "the lecturer who doesn't understand how to set up the required test". "Although formal training may be provided for the lecturers on how to set up a site with all the tools available to them, how the tool works, and what the correct method of using the tools is, how to get the most out of the course tools for teaching purposes and understand the value of the tool for teaching and assessment purposes, we don't have enough academics/administrators who fully understand the tool merely because they fail to attend training sessions provided", added S2(16). Another respondent stated that "training courses are one aspect, but experience counts a whole lot more" S2(28).

Resistance to change – A notable resistance faced by participants came from their own colleagues and faculty members, "other academics, generally within the department – mainly from courses who don't use it, and academics who don't feel comfortable using Information Systems/Technology" stated S2(53). He added that "they would rather use the old-fashioned paper method. They just resist change in whatever form it is. Any change is bad, let's keep it to how we were in 1920 – whatever year they were born. They feel intimidated by the use of technology. They don't seem to bother much what the students think; it's more about how they feel – so if they are more comfortable using traditional assessment then the students will have to do it that way". S2(16) added that "they fear technical mishaps; what if the power goes down (in the past 5 years it has never gone down), or what happens if the computer freezes or it's not a good experience for the students". Student resistance may also occur, because they feel "there is too much work in the course" compared to the courses that use traditional assessment" stated S2(12).

Security was an issue indicated by only one respondent, S2(46), who had faced a situation "where answers were inadvertently displayed during the test".

Test bank limitations – Formative assessment can be problematic if the database is not large enough to create multiple versions of tests, especially for large classes.

5.2.5 Conclusion and summary of Study 2

Study 2 was an interview-based study which addressed similar issues as Study 1 on the usage of e-assessment, but in greater detail and with double the participants, all of whom were adopters of e-assessment. The findings cannot be statistically compared, because there was an overlap in the samples. But it is notable in Study 2, more than a year after Study 1, that more than half of the participants had adopted e-assessment for only one to two years. This high occurrence of new adopters is evidence that the usage of e-assessment is on the increase in South African tertiary institutions. The quantitative component was discussed in Section 5.2.3, while the purely qualitative concepts were discussed in Section 5.2.4. As in Study 1, manual analysis and coding was undertaken to extract themes and patterns from the textual interview transcripts. Benefits associated with e-assessment, addressed in Section 5.2.4.1, included automated feedback, consistency, rapid scoring, reduced uncertainty, assessment frequency, question sharing and reuse, and student engagement; all of which can support assessment practices in both blended learning and open distance learning, where large numbers of students are widely dispersed. Despite disadvantages mentioned in the literature and articulated by certain users and non-users of e-assessment in Section 5.2.4.2, the benefits outweigh the drawbacks, particularly in a milieu of rapidly increasing student numbers.

Furthermore, educators' workloads can be reduced due to the use of automated assessment in place of some of the manual marking.

From the student's perspective, the use of e-assessment as formative self-assessment can provide opportunities for practice of skills and consolidation learning in personal time and space. In Study 2 the use of formative assessment was more than double the use of summative assessment – indicating valuable use of e-assessment for practice and revision.

Finally, the researcher acknowledges that e-assessment tools must be used in appropriate contexts and in conjunction with other forms of traditional assessment for the holistic assessment of a student's performance and knowledge.

5.3 Study 3

Study 3 was a survey-based study where participants were all adopters of e-assessment. Data gathered during this study was predominantly quantitative in nature. Hence the quantitative data analysis for this study is presented in Section 5.3.3. Although the respondents included some international academics, their responses were excluded from the results reported here, as they fell beyond the scope of South African tertiary institutions.

5.3.1 Introduction to Study 3

In Studies 1 and 2, the researcher studied the extent and nature of usage of electronic assessment tools within Computing-related academic departments and schools at South African tertiary institutions, as well as users' satisfaction with the tools (Singh & de Villiers, 2010). In their role as the baseline studies to establish the status quo in the early stages of this longitudinal research, Study 1 and Study 2 helped to refine the actual research project and contribute to the design of research instruments for Study 3. Studies 1 and 2 aimed to establish a general context for subsequent research by determining the situation during the period from late 2009 (Study 1) to 2011 (Study 2 in 2010/2011).

The interviews in Study 2 contributed to the identification of features that users required in e-assessment tools, and hence to the identification criteria for evaluating e-assessment systems. These were compiled and incorporated in Table 5.35 which was used in the evaluation of the SEAT and e-SEAT Frameworks.

Study 3, as presented in Table 5.29 (similar to Table 4.7 in Section 4.5.3.), was undertaken later in 2011, and investigated the different types of MCQs (Section 3.1.5) adopted and their relevance to higher order thinking skills (HOTS) (Section 3.1.3). Specifically, the investigation of the varying MCQ types adopted by participants, served to confirm the items identified in the literature, which are included in the Question Types category of the SEAT Framework.

Study 3 aimed to establish answers to Research Questions 2 and 3, as given in Chapter 1 and Chapter 4:

Research Question 2

What types of questions are being adopted in e-assessment systems in South Africa?

Research Question 3

How appropriate are these questions (identified in Research Question 2) for testing higher order thinking skills (HOTS)?

Study 3 also investigated two subsidiary questions:

- For which levels of study are the types suited?
- What benefits are associated with e-assessment?

Most of the material in Sections 5.3.2 and 5.3.3 is based on an article (Singh & de Villiers, 2012) published in a Progressio Journal of 2012 (Appendix L2). The research presented in the publication was conducted by the researcher as an integral part of her PhD studies and was also used for the journal article.

Table 5.29: Summary of Study 3 as outlined in Figure 5.1

Study 3 (March – June 2011) – MCQ adoption patterns and HOTS	
Respondents	64 respondents from 15 South African Institutions (92 in total including international participants)
Data Collection	Survey Instrument – Questionnaire 2 (See Appendix E)
Data Analysis	Quantitative Analysis – basic statistical analysis
Purpose	To obtain information on the different types of MCQs adopted To understand how applicable these types of MCQs are to higher order thinking skills (HOTS).

5.3.2 Questionnaire participants

By means of a questionnaire, MCQs were investigated further in Study 3 (the present study), as shown in Figure 5.1.

Questionnaire participants were a combination of Study 1 and Study 2 respondents and referrals, both local and international, that is, the sample was a combination of purposive sampling and snowball sampling (Cohen et al., 2011). Questionnaires were distributed via email to 132 potential participants, of whom 92 responded. Sixty-four (64) participants were from South Africa, representing fifteen institutions, as outlined in Table 5.30. The largest numbers, thirteen each, were from UCT and UFS; followed by WITS with eleven; UP five; and UNISA four. Eighteen (18) of the original 92 questionnaires were excluded, because their data was inadequate. Ten (10) of the 92 questionnaires were from international respondents, hence were also excluded, resulting in a final number of 64 participants whose data was used.

Table 5.30: Participant distribution by institution

Institution	Frequency	Percentage
UFS	13	20.3
UCT	13	20.3
WITS	11	17.1
UP	5	7.8
UNISA	4	6.3
NWU	3	4.7
SUN	3	4.7
UKZN	2	3.1
UWC	2	3.1
NMMU	2	3.1
WSU	2	3.1
DUT	1	1.6
CUT	1	1.6
Monash	1	1.6
UL	1	1.6
TOTALS	***64***	***100***

UCT = University of Cape Town	**UFS = University of Free State**	**UKZN = University of KwaZulu-Natal**	**UNISA = University of South Africa**	**UP = University of Pretoria**	**WITS = University of Witwatersrand**	**CUT = Central University of Technology**
NWU = North West University	**SUN = Stellenbosch University**	**UWC = University of Western Cape**	**NMMU = Nelson Mandela Metropolitan University**	**WSU = Walter Sisulu University**	**DUT = Durban University of Technology**	**UL = University of Limpopo**

The research was initially aimed at Computing users, but due to high usage and earlier adoption of e-assessment tools, non-Computing users were incorporated. The analysis in this section shows composite findings, as well as some results by user type. Computing users, as illustrated in Figure 5.12, 21.1% came from UCT and 18.4% from UFS, while 10.5% came from UNISA and UP, while with non-Computing users, as shown in Figure 5.13, 30.8% were from WITS, followed by UFS (23.1%) and UCT (19.2%). The Computing users were distributed over fourteen institutions, while non-Computing users came from eight. UCT and UFS, with the most respondents overall, had relatively high numbers in both groups.

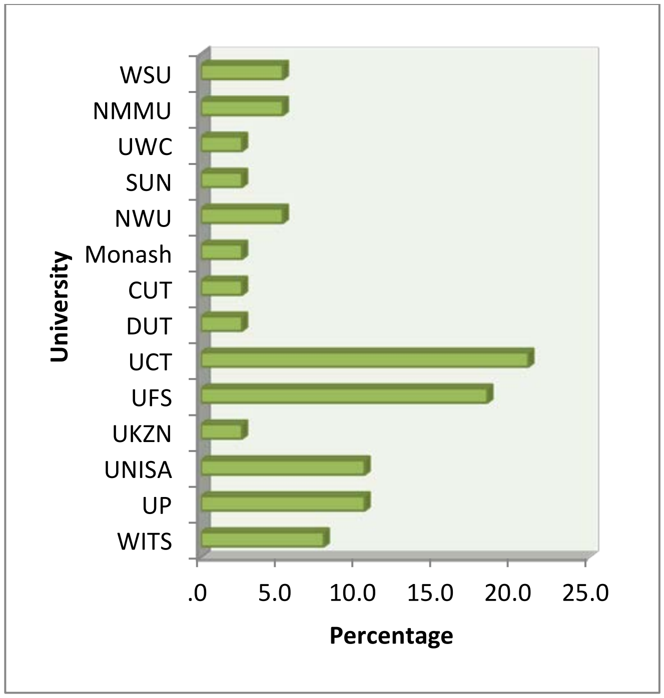


Figure 5.12: Computing participants per Institution

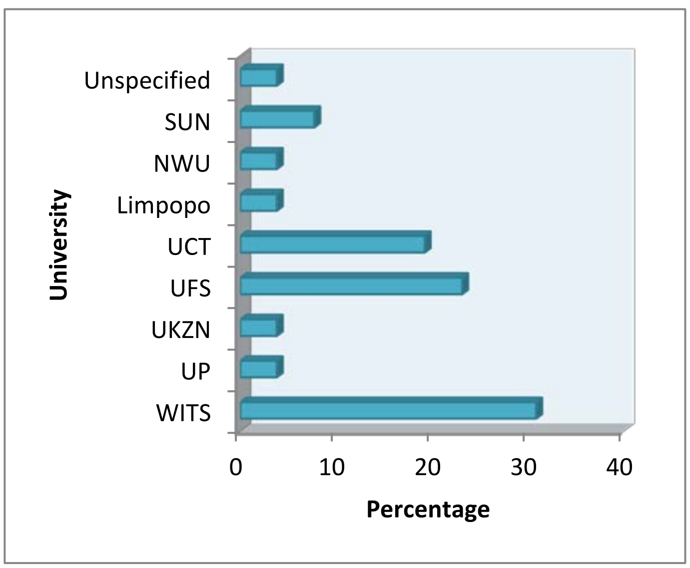


Figure 5.13: Non-Computing participants per Institution

Figure 5.14, depicts eleven faculties, with the greatest participation coming from Faculties of Science, Engineering and Technology (28.1%), Commerce and Economic Sciences, and Management and Law (17.2%), and Higher Education (12.5%). Health-related departments tend to be early adopters of educational technology, often due to students' practical work, limiting their time in classrooms. Health Science, and Natural and Agricultural Sciences, each accounted for 10.99% of the participation.

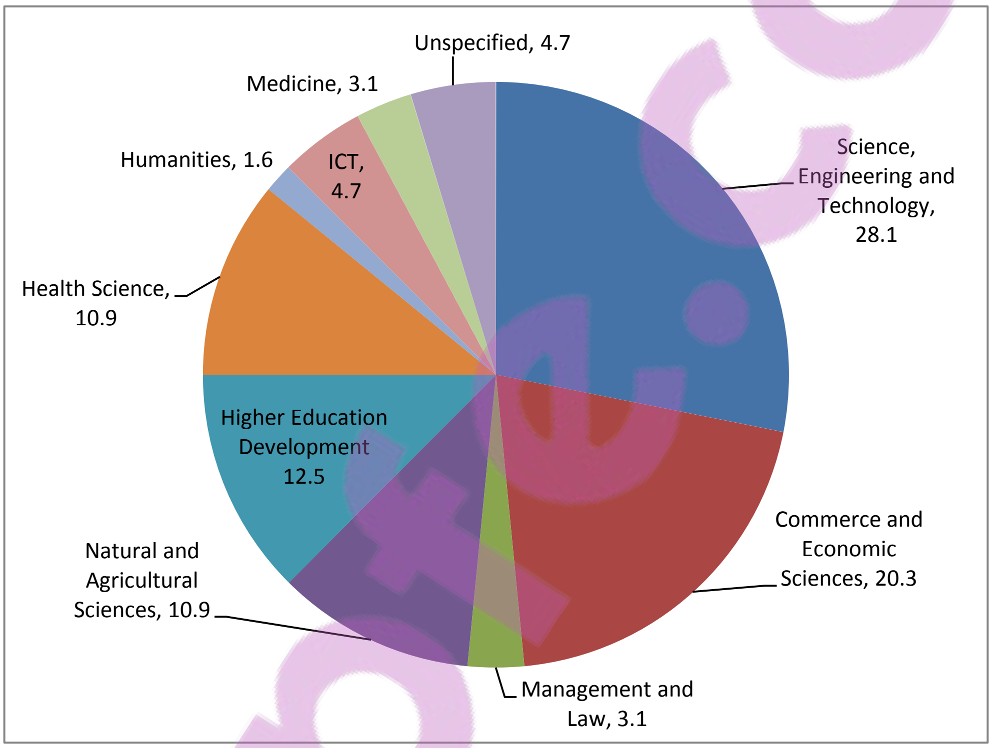


Figure 5.14: Faculty participation

As presented in Table 5.31, participants represented twelve schools/departments, with a high number from schools or departments of Computer Science (28.1%), Information Systems (14.1%) and Information Technology (9.4%). There was readiness in the Education discipline to adopt technology in departments/sections of Educational Technology (9.4%). Mathematics (7.8%) and Family Medicine (6.3%) also had high participation.

Table 5.31: Participant distribution by school/department

School/Department	Frequency	Percent
CS	18	28.1
Information Systems (IS)	9	14.1
IT	6	9.4
Educational Technology	6	9.4
Mathematics	5	7.4
Family Medicine	4	6.3
Psychology	1	1.6
ICT	1	1.6
Economics	1	1.6
Information Sciences	1	1.6
Anatomy	1	1.6
Business Management	1	1.6
Therapeutic Sciences	1	1.6
Other	9	14.1
TOTAL	***64***	***100***

Of the 64 participants whose questionnaires were used, 57.9% were senior lecturers or lecturers, as outlined in Table 5.32. One participant failed to indicate his/her position within the department. Some participants (34.6% of non-Computing users and 8% of Computing users) had support staff to administer assessments, while certain Computing users – who were comfortable implementing technology managed e-assessment personally.

Table 5.32: Participant distribution by position held

Position	Frequency	Percentage
Professors	10	15.6
Senior lecturers	12	18.8
Lecturers	25	39.1
Tutors	1	1.6
Support Staff	15	23.4
Total	***63***	***98.5***

5.3.3 Findings of Study 3

To answer Research Question 2, an understanding of the usage of the various types of questions in the MCQ genre, as outlined in Section 3.1.5, was undertaken. Questionnaire 2 used in Study 3, is provided in Appendix E.

Sixteen types of questions in the MCQ genre were listed in the questionnaire, which were gathered from the literature study presented in Section 3.1.5. The participants were asked to indicate all the e-assessment types they had used over the years and across modules.

The most common question types adopted by the 64 South African academics, Computing and non-Computing together, were Multiple Choice: single-response (40.90%) and Multiple Choice: multiple-response (17.44%), followed by True/False (14.45%), Simulations (10.44%), and the others below 10%. Text-Input/Short-Answer had a 14.55% response, but in all instances, these were marked manually.

Participants were also asked to indicate adoption of the various question types in intervals:

0%, 1-29%, 30-69%, 70-99% and 100%. For example, if a participant selects 30-69% for True/False, it indicates that the participant adopts True/False questions 30-69% of the time in his/her assessments. The adoption patterns of Computing and Non-Computing users, according to this classification, are presented in Table 5.33.

Table 5.33: Average adoption of question types – comparison of groups

Question Types	Computing (%)	Non-Computing (%)
Multiple choice: single response	40.18	41.92
Multiple choice: multiple response	11.99	25.19
True/False	13.61	15.65
True/False with explanation	4.18	4.04
Fill-in-the-Blanks/Completion	9.72	5.37
Simulation	14.68	4.42
Matching Items	5.39	9.19
Extended-matching items	5.39	9.19
Selection/Drop-down lists	2.84	4.04
Ranking	5.14	0.58
Diagram/Video Clips	7.14	5.37
Drag-and-Drop	1,62	2.48
Reordering/Rearrangement/Sequencing	1.62	1.15
Categorising	1.62	0.58
Hotspots	1.22	6.71
Text-Input (Short-Answer)	15.05	6.09

Chi-square goodness-of-fit tests were performed on this data to see whether, for each question type, there was significant selection of a specific usage percentage. Despite the indication in Table 5.33 that Computing users seem more willing to adopt a variety of types, there were also significantly more respondents than expected who indicated a 0% usage of the non-standard question types, that is, no use of types outside direct multiple choice and true/false. In each case $p<.0005$.

Various cross tabulations were done, one of which, Usage/Faculty indicated a significant relationship, in that significantly more than expected respondents from Management Sciences (Fisher's exact (N=32) = 10.252, p=.024) were using Ranking questions in up to 30% of their assessments.

The second question in Questionnaire 2, aimed to answer Research Question 3; thus it investigated the usefulness of the different MCQ types in assessing higher order thinking skills (HOTS).

Bloom's taxonomy (Bloom, 1956) presents a progression in levels of thinking, starting at the concrete lowest order, Level 1, with *facts*. Thereafter, students *comprehend* meanings and implications of the facts, Level 2. In Level 3 they *apply* their learning, which helps them to solve problems and transfer knowledge to related situations. In *analysis,* Level 4, students can classify, categorise, discriminate and detect information, as well as compare and contrast concepts. *Synthesis,* Level 5, involves combining ideas, planning, forming solutions, and creating new information. *Evaluation* on Level 6 requires taking decisions, ranking concepts and making judgments regarding information and situations (Bloom, 1956; Passey, 2011).

Participants were asked, in terms of relevance to HOTS, to rate the types, in four categories: <Not useful>, <Undecided>, <Useful>, <NA/unfamiliar>. Computing and non-Computing users evidenced very similar patterns, shown in Figures 5.15 and 5.16 respectively.

Figure 5.15 illustrates the non-Computing users' opinions on the relevance of each MCQ question type and its applicability to HOTS.

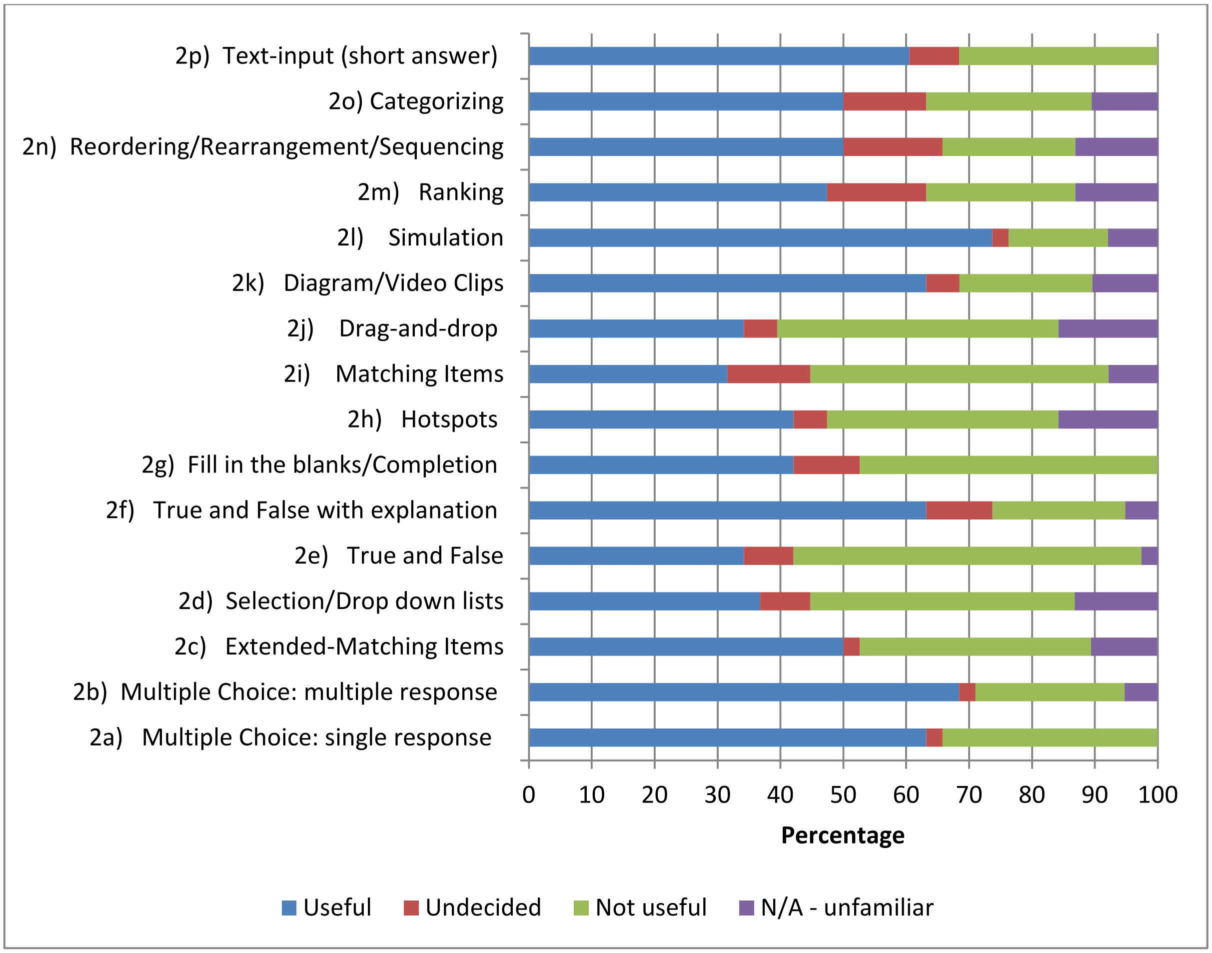


Figure 5.15: Usefulness of question types for HOTS – Computing users

Figure 5.16 presents Computing users' opinions on the relevance of the varying MCQ question types and their applicability to HOTS.

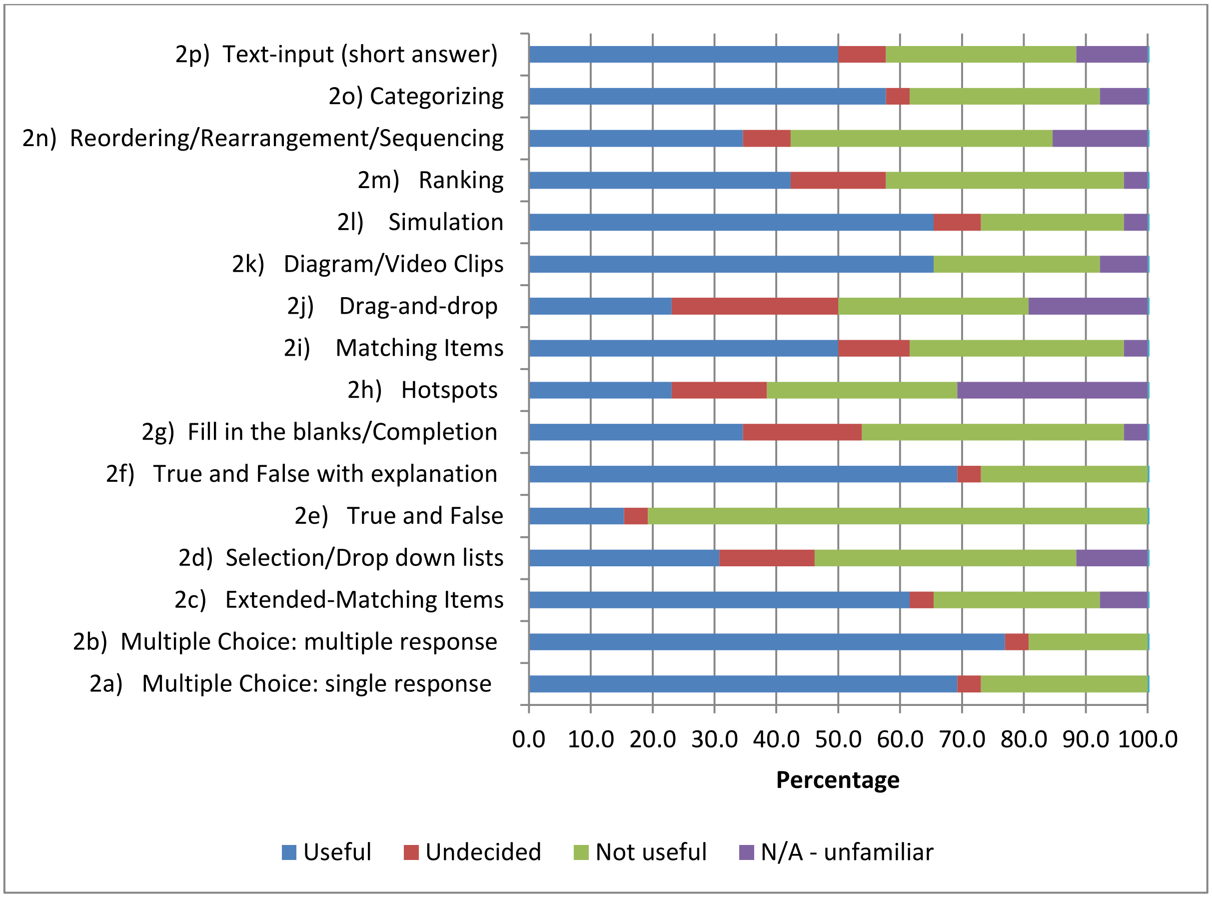


Figure 5.16: Usefulness of question types for HOTS – Non-Computing users

Chi-square goodness-of-fit analysis was done to ascertain which responses were selected significantly more often than the others. Significance is taken for any p-value less than 0.05, indicating 95% confidence. The lower the p-value, the more significant is the result.

- <Useful> was chosen by significantly more respondents than expected for Extended-Matching Items; True/False with explanation; Diagram/Video Clips; Simulation; Multiple Choice: single response; and Multiple Choice: multiple response. For these six types, significance was p <.0005.
- Ranking (p=.001); Reordering/Rearrangement/Sequencing (p=.001); and Categorising (p=.002) were significant to levels indicating high usefulness for assessing HOTS.

- <Not useful> was chosen significantly more often for True/False (p<.0005); Matching Items (p=.002); and Drag-and-drop (p<.017).

Selection for the response options was not even across options. Multiple choice: single response was the only question type with which all respondents were familiar. The above analysis applies to both Computing (Figure 5.15) and Non-Computing users (Figure 5.16).

The third question investigated which question types the interviewees felt were useful for the four levels of study. The findings are consolidated in Table 5.34 below.

Table 5.34: Suitability of questions per level of study

(Comp = Computing; Non-Comp = Non-Computing)

Question Types	Year 1 % of Interviewees		Year 2 % of Interviewees		Year 3 % of Interviewees		Year 4 % of Interviewees	
	Comp	Non-Comp	Comp	Non-Comp	Comp	Non-Comp	Comp	Non-Comp
MCQ: Single Response	92.1%	96.2%						
True/False	76.3%	80.8%						
EMI	73.7%	57.7%						50.0%
Selection/Drop Down Lists	73.7%	69.2%						
MCQ: Multiple Response	81.6%	69.2%	63.2%	69.2%				57.7%
Fill-in-the-Blank/ Completion	78.9%	69.2%						
Hotspots	52.6%		52.6%					
Matching Items	68.4%							
Drag-and-drop	68.4%							
True/False with explanation	60.5%		65.8%	73.1%				50.0%
Diagram/Video Clips	60.5%		57.9%	65.4%	63.2%	65.4%	52.6%	-
Simulation	68.4%		65.8%		63.2%	61.4%	50.0%	61.5%
Ranking	68.4%							
Reordering/ Rearrangement/ Sequencing	68.4%							
Categorising	57.9%		57.9%					

Table 5.34 shows that, for first-year level (Year 1), there was broad use of types, but particularly MCQ (single response) and True/False. Computing academics explored every type, especially MCQ: multiple response, Fill-in-the blank and EMI questions. On second level (Year 2), users were selective

and made greater use of the more advanced types, while at third level (Year 3), usage was concentrated on Diagram/Video Clips and Simulation. These question types frequently implement Bloom's analysis and synthesis categories. There was a similar tendency at Honours (Year 4), where adoption was greatest on Diagram/Video Clips and MCQ (multiple response), which involve Bloom's evaluation category.

5.3.4 Conclusion and summary of Study 3

The study has established that the usage by South African academics of the various types of multiple choice questions is concentrated on Multiple-Choice: single and multiple response, True/False and Fill-in-the Blank questions. The more novel question types such as Extended-Matching Items, True/False with explanation, Diagram/Video Clips, Simulation, and Multiple Choice: single and multiple response, were found by users to be relevant for assessing higher order thinking skills and upper levels of Bloom's Taxonomy. A notable finding (Table 5.34) was the low usage of the non-standard types of MCQs. Multiple choice: single response was extensively used, while there was moderate use among Computing users of multiple choice: multiple response, true/false, fill in the blank, and simulation questions. This highlights the need to promote other question types.

For first-year students, adoption is concentrated on MCQ (single response) and True/False; while at second, third and honors levels, Diagram/Video Clips and Simulation are the most used.

A further consequence of Study 3 was acknowledgement of the importance of including these various types and formats of MCQs in the SEAT Framework.

5.4 Framework evolution

As illustrated in Figure 3.16 in Chapter 3, the development of the SEAT Framework was evolutionary. Table 3.1 outlined the 91 criteria for inclusion in the framework that were compiled from the literature studies presented in Chapters 2 and 3. Subsequent to Studies 1 to 3, the list of 42 additional criteria identified from the interviews conducted, which were deemed necessary to include in the SEAT Framework, are summarised in Table 5.35. These 42 additional criteria answered Research Question 4 – by identifying important evaluation criteria for e-assessment tools, specifically in the South African Higher Education context.

Research Question 4	***What are the requirements for selecting or personally developing an electronic assessment tool?*** ***Practice: What criteria are used in practice in South African higher education institutions for the selection and use of electronic/online testing and assessment tools?***

These criteria were gathered from the qualitative comments made by participants in their verbal discussions during the personal and telephonic interviews in Study 2, as well as from Questionnaire 2 during Study 3. The researcher classified the additional criteria identified in these empirical studies into the eleven categories presented in Table 3.1. An additional category, namely, the Question Types category, was created and incorporated in Table 5.35, following Study 3, which identified sixteen varieties of MCQ question types, as essential for inclusion in an e-assessment system. No additional criteria were identified during Studies 1 to 3 for three categories, namely, *Test Bank, Import/Export Data, and Training categories.*

Table 5.35: Framework version 1b - e-assessment criteria identified in the empirical studies (Ref: Personal and telephonic interviews, Questionnaire 2)

CATEGORY	CRITERIA
Interface Design	1. Students can view all tests available to them (Interviews) 2. Students see times and locations of formal examinations (Interviews) 3. Multiple windows can be provided (Interviews) 4. Display marks for each question and section (Interviews)
Question Editing	1. Can author original questions (Interviews) 2. Question metadata can be incorporated (Interviews) 3. Offline question creation is possible (Interviews) 4. Spell checker is incorporated (Interviews) 5. Ability to approve or disapprove a question and add comments (Interviews)
Assessment Strategy	1. Students can sit a test as many times as they like for self-assessments (Interviews) 2. Can access a test without authentication for self-assessments (Interviews)
Test and Response Analysis/Reports	1. Students can comment on questions and tests (Interviews) 2. Academics can view feedback on tests and questions from students (Interviews) 3. Markers can add notes about students relating to their responses (Interviews) 4. Automated cheating spotter facility is incorporated (Interviews) 5. Entire test is viewable as it was completed by the student (Interviews)
Test Bank	None
Security	1. System must be robust and not result in crashes on the server or browser (Interviews) 2. All user actions are logged – student (where they sat, which IP address), marker (which question marked, when marked) (Interviews) 3. Already answered questions cannot be altered (Interviews) 4. Can enter details of students who cheat (Interviews)
Compatibility	1. Results can be exported to spreadsheets or statistical analysis software (Interviews) 2. Can enter marks, dates of submission and other details of non-CAA tests (Interviews)
Import/Export Data	None
Ease of Use	1. Can add details of room numbers and invigilators (Interviews) 2. Can access details of students sitting a test at a particular time (Interviews)
Technical Support	1. Resilient network required (Interviews) 2. New functionality can be incorporated without reinstalling the system, (Interviews)
Training	None

Question Types (new category created)	1. Multiple choice: single response (Questionnaire 2, Interviews) 2. Multiple choice: multiple response (Questionnaire 2, Interviews) 3. True/false (Questionnaire 2, Interviews) 4. True/false with explanation (Questionnaire 2) 5. Fill-in-the-Blanks/Completion (Questionnaire 2, Interviews) 6. Simulation (Questionnaire 2) 7. Matching Items (Questionnaire 2) 8. Extended-matching items (Questionnaire 2, Interviews) 9. Selection/Drop down lists (Questionnaire 2) 10. Ranking (Questionnaire 2) 11. Diagram/Video Clips (Questionnaire 2) 12. Drag-and-Drop (Questionnaire 2) 13. Reordering/Rearrangement/Sequencing (Questionnaire 2) 14. Categorising (Questionnaire 2) 15. Hotspots (Questionnaire 2, Interviews) 16. Text Input (short answer – which would be marked manually) (Questionnaire 2)

The 91 criteria from the literature (Table 3.1) and the 42 criteria from the interviews and Questionnaire 2 (Table 5.35) were merged into a composite framework with 133 criteria, subdivided into twelve categories. Some of the criteria identified, which were regarded as compound criteria (for example Criterion 4 in Category 1, Table 5.35, which enquires about marks per question and marks per section in one criterion), were further subdivided to facilitate more accurate data collection. Thus, this composite framework, which was entitled the Pilot Framework (Appendix F1), consisted of 147 criteria (as a net result of the subdivision and combination of 15 criteria, see Table 7.2). This Pilot Framework would serve as the initial version of SEAT to be evaluated and refined through the action research series conducted from Study 4a to Study 6.

5.5 Chapter conclusion

Phase 1 of this study, as depicted in Figure 5.1, set the foundation for the Evaluation Framework (SEAT and subsequently e-SEAT) that was developed in Phase 2. Phase 1 comprised 3 studies, namely, Studies 1 to 3 to establish the nature, context and extent of adoption of e-assessment, particularly with Computing-related disciplines at South African Higher Education Institutions. Phase 1 provided both quantitative and qualitative data. Sections 5.1, 5.2 and 5.3 discussed Studies 1, 2 and 3 respectively. A description of the evolution of the SEAT Framework was presented in Section 5.4 along with Table 5.35, which listed 43 additional criteria identified in Phase 1 to be incorporated in the prototype evaluation framework, SEAT.

This chapter presented the data analysis of the three studies undertaken in Phase 1 of this action research, namely, Studies 1 to 3. The purpose of these studies was to set the context for future research on the requirements for, and evaluation of, e-assessment systems and tools, which is described and delivered in Phase 2 of this research, presented in Chapter 6. The action research process undertaken in Studies 4, 5 and 6 in Phase 2, resulted in the evolution of SEAT to e-SEAT, through the iterative evaluation and development of each version, until the e-SEAT Final Framework was created, as illustrated in Figure 6.2.

The outcome of Chapter 5 was that Research Questions 1, 2, 3 and 4 were answered. Question 1 asks 'What is the extent and nature of use of electronic assessment in Computing-related departments in South African universities? The findings in response to this provided the context of adoption of e-assessment tools within Computing-related academic departments/schools at South African higher education institutions.

Question 2 was 'What types of questions are being adopted in e-assessment systems in South Africa?' Participants' responses to this contributed strongly to the SEAT Framework, causing the researcher to incorporate an additional category called 'Question Types'.

Question 3 was 'How appropriate are these questions (identified in Research Question 2) for testing higher order thinking skills (HOTS)?' Despite the potential of MCQs in assessing HOTS, the reponses identified low usage of HOTS for this purpose.

Question 4 was 'What are the requirements for selecting or personally developing an electronic assessment tool?' The data collected to answer this question answered the second part of Research Question 2 question by identifying criteria used in practice in South Africa for the selection and use of electronic/online testing and assessment tools. This generated the major deliverable of the chapter, namely the structured set of criteria for the SEAT Framework, gathered from qualitative comments in the empirical research, and presented in Table 5.35. These were merged with the initial SEAT Framework, garnered from the literature (Table 3.1) and incorporated in the evolving SEAT Framework which served as the input to the evaluative series of sub-studies in Study 4.

CHAPTER 6 Data presentation and analysis of Phase 2 Studies

Phase 1, incorporating Studies 1 to 3, set the context for future research on the requirements for, and evaluation of, e-assessment systems and tools. Phase 2 of this action research comprised three studies, namely Studies 4 to 6, as depicted in Figure 6.1, which is the second section of Figure 4.8.

This phase commenced with the design and development of the initial version of the SEAT (Selecting and Evaluating an e-Assessment Tool) Framework, based on two sets of data, namely, primary data obtained from participants in Studies 1 to 3, as presented in Table 5.35 and secondary data from the literature categorised in Table 3.1 in Section 3.2. These two forms of data, integrated into a single set, comprised the input into Study 4, where SEAT was refined. Study 4 was iterative in nature, with four substudies, progressing from a Pilot Study (Section 6.1.1) through an Evaluation Study (Section 6.1.2), to a Proof of Concept Study (Section 6.1.3), culminating in an Application Study (Section 6.1.4). The SEAT Framework was validated through the Application Study, which made possible the transition from the manual SEAT Framework to an electronic version named e-SEAT (electronically Selecting and Evaluating an e-Assessment Tool).

Following the development of e-SEAT, it in turn needed to be evaluated, applied and validated. Study 5 (Section 6.2) and Study 6 (Section 6.3) thus saw participants evaluating, applying and validating the electronic framework developed.

The development of SEAT into e-SEAT was an iterative and evolutionary process as is the case with action research studies. Each of the six successive versions of SEAT/e-SEAT is named according to the respective study for which it serves as input, for example, the 'Pilot Framework' was the input for the 'Pilot Study' and so on.

Figure 6.2 shows the progressive development of the framework through the first four substudies on the paper-based manually-operated SEAT. The framework progressed from the Pilot Framework, through the Evaluation Framework, to the Proof of Concept Framework, and culminated in the Application Framework. The conversion of SEAT from a manual framework to an electronic version named e-SEAT, occurred after Study 4. This electronic framework was evaluated, applied and validated by participants during Studies 5 and 6 as it evolved through the Evaluation Framework and the Validation Framework, respectively, until the Final e-SEAT Framework was reached. The electronic versions are shown on the right hand side of Figure 6.1. Following the e-SEAT validation study, e-SEAT was considered to be the ultimate product.

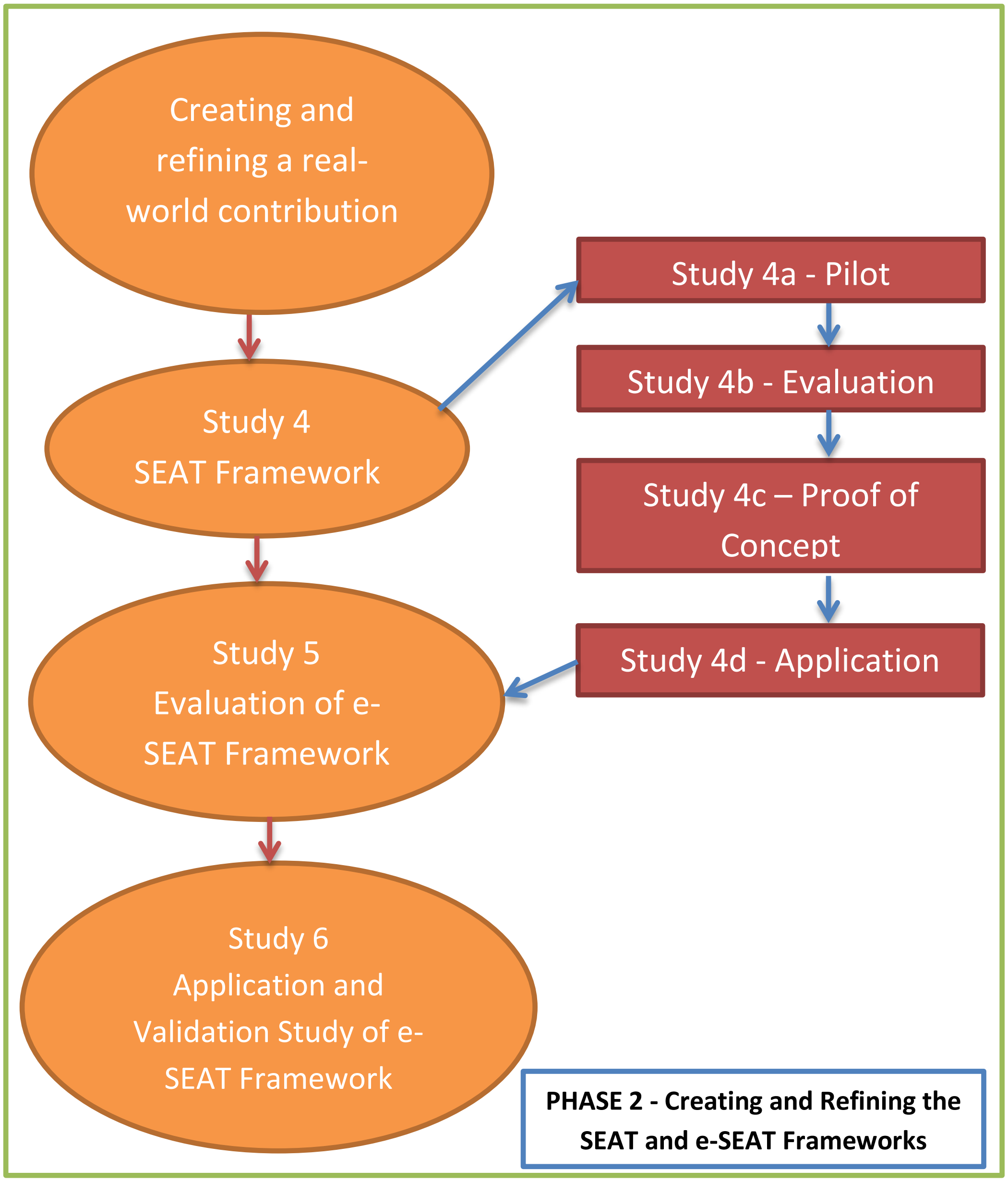


Figure 6.1: Phase 2 of this research

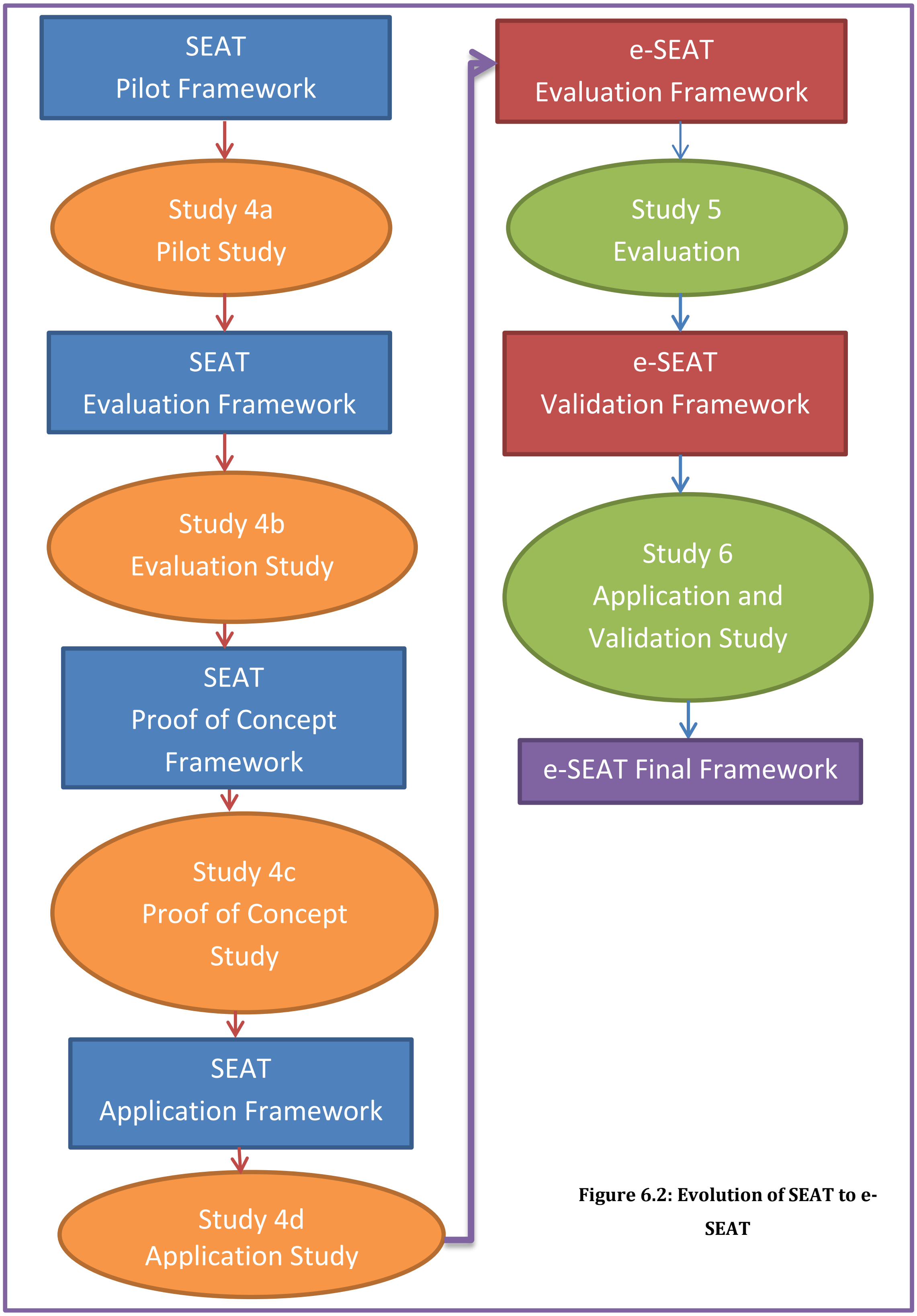


Figure 6.2: Evolution of SEAT to e-SEAT

6.1 Study 4

Study 4 aimed at the initial creation and refinement of a comprehensive framework to evaluate e-assessment systems. The SEAT Evaluation Framework was a combination of the criteria identified in the Literature Study presented in Chapter 3 (see Table 3.1) by participants in the empirical studies in Phase 1 (see Table 5.35). To achieve this, it consisted of four substudies, each of which moved a step closer to the final SEAT Framework. The Pilot Study, Study 4a (Section 6.1.1) was used to obtain essential feedback from two participants on the design, content and validity of the first-draft version of the merged framework. Study 4b, the Evaluation Study, (Section 6.1.2) determined which of the criteria are essential for any e-assessment tool. Study 4c (Section 6.1.3) provided insight into the essential criteria to be included in the SEAT (Selecting and Evaluating an e-Assessment Tool) Framework. Finally, Study 4d (Section 6.1.4) took the SEAT Framework to participants so they could apply it to an existing e-assessment system. The four versions of the SEAT Framework are provided in Appendices F1, G, H1 and I respectively; while the e-SEAT Final Framework can be viewed in Appendix J1.

Research Question 5 is answered in Sections 6.1.1, 6.1.2 and 6.1.3; while Research Question 6 is partially answered in Section 6.1.4.

Research Question 5	***What categories and criteria should be incorporated in a prototype framework to evaluate electronic assessment systems?***

Research Question 6	***How appropriate and effective is the proposed framework?***

6.1.1 Study 4a – Pilot Study

Table 6.1: Summary of Study 4a as outlined in Figure 6.1

Study 4a (April 2012) – Pilot Study	
Respondents	2 Participants from UKZN
Data Collection	SEAT Pilot Instrument (See Appendix F2) and Questionnaire (See Appendix F3)
Data Analysis	Qualitative – content and discourse analysis
Purpose	To obtain initial critical feedback on the design, content and validity of the instrument to be used in the data collection process of the main Studies in this research

Study 4a, initially presented in Table 4.8 in Section 4.5.4, is summarised in Table 6.1. It contributed to answering Research Question 5. It commenced with a preliminary identification by the researcher, of pertinent criteria that should be included in the framework being developed to evaluate e-assessment tools. Ninety one (91) criteria were gleaned from the literature reviewed (see Table 3.1), while another 42 criteria were generated from points that emerged from the surveys, interviews and observations undertaken in Studies 1 to 3 (see Table 5.35). The criteria thus included those identified locally in South Africa, specifically related to the South African context, as well as information from the large body of literature surveyed from the international and local arenas. It had become evident that various e-assessment tools provide multiple functionalities, far more than the features that most users are accustomed to in basic systems. Hence, the framework included features and facilities that are context-dependent and may not be required in every system and under all circumstances.

The criteria groupings were based on the features they related to in the e-assessment tool. The composite list of criteria from the two sources above, were subsequently categorised into the following ten categories:

1. Interface Design
2. Question Editing
3. Assessment Strategy
4. Test and Response Analysis/Reports
5. Test Bank
6. Security
7. Compatibility
8. Ease of Use
9. Technical Support
10. Question Types

The *Import/Export Data* and *Training* categories given in Table 3.1 were merged into the categories of *Compatibility* and *Ease of Use*, respectively, as presented in the Pilot Framework (Appendix F2). As explained in Section 5.4, the *Question Types* category was created for inclusion in the SEAT Framework following Study 3. Furthermore compounded criteria were further subdivided to facilitate more accurate data collection resulting in 147 criteria in ten categories (as a net result of the subdivision and combination of 15 criteria, see Table 7.2).

The initial framework termed the Pilot Framework was developed from the Word document, with 147 criteria, (see Appendix F1) and hosted on the online survey tool *Survey Monkey* (see Appendix F2). During the process of converting the Word version of the Pilot Framework to the version hosted on *Survey Monkey* a further 15 criteria were identified for subdivision. Thus the version distributed to participants contained 162 criteria (see Table 7.2 and Appendix F2) in the same ten categories listed above. The use of *Survey Monkey* facilitated easier instrument distribution and data analysis. An invitation to participate in the Pilot Study (Figure 6.3) was sent to three colleagues of the researcher, who had not been part of Studies 1 to 3, asking them to serve as evaluators to pilot test the framework. They were selected by the researcher, as highly meticulous and critical academics. One invitee declined to take part in this study, resulting in two participants serving as evaluators.

Dear xxx,

I have a request. I have completed the final instrument for my PhD study data collection. I would like some colleagues, who are not part of the study, to pilot the initial version of the framework developed.

As the study is designed to produce an evaluation framework, it is a little lengthy and would require approximately 30 minutes to complete, but it has been designed as an online survey with rating questions to speed up the process, with spaces for comments as required. I have selected you, as a meticulous and critical colleague (meant in a positive way), and your input would be greatly appreciated to improve the survey to obtain the best possible data …

Figure 6.3: Extract from the invitation to participate in the Pilot Study

The pilot framework, in its initial form, is given in Appendix F2. As explained in Figure 6.3, the 162 criteria (Table 7.2) in this pilot framework were presented as rating questions to expedite the process, with spaces for comments as required. The participants were informed that, since the study was designed towards developing a comprehensive framework, it was rather lengthy and required approximately 30 minutes to complete.

The aim of the Pilot Study was for the participants to rate the criteria according to the Likert scale ratings presented, to ascertain which criteria were most relevant specifically for the South African context. For each of the criteria listed in each category, the participants were required to rate the criterion on a scale of 1 to 7, where 1 = Extremely Important and 7 = Not at all Important. Furthermore, at the end of each category, participants were requested to provide:

- General comments pertaining to the category, and
- Any comments that they thought were essential, but not already included in the category.

The Pilot Study took place over seven days due to the rigorous review undertaken by the participants. Both provided detailed critiques of the Pilot Framework. Their comprehensive and meticulous input helped to improve the survey content, so that the best possible data could be obtained in the next stage, namely the main Evaluation Study.

Individual comments on the criteria and the categories included suggestions on:

- rewording actual criteria to improve clarity,
- moving criteria to other categories where they would fit better,
- explanation of terms used within the criteria, expansion of acronyms, and in some instances providing examples of what the criterion was intended to evaluate,
- identification of similar criteria across and within categories, which could be regarded as duplication and should be eliminated,
- additional criteria that were necessary but not already included, and
- rephrasing, deletion and addition of categories.

In addition to commenting on the individual criteria in each category, the participants were asked a few questions at the end of the Pilot Study. The themes of these questions are summarised in Table 6.2, also indicating the sections where they are addressed.

Table 6.2: Question themes for pilot participants

	Themes	Section where addressed
1	Instruction clarity	6.1.1.1
2	Criteria ambiguity	6.1.1.2
3	Understandability	6.1.1.3
4	Time saving	6.1.1.4
5	Timing/Duration	6.1.1.4 and 6.1.1.5
6	Ways to Improve	6.1.1.6

6.1.1.1 How clear were the instructions?

PP1 (Pilot Participant 1) found the instructions to be clear, however PP2 (Pilot Participant 2) felt that certain instructions needed clarification.

6.1.1.2 Were any of the evaluation statements unclear or ambiguous? If so, please specify which, and explain why.

Most of the changes that occurred in the Pilot Framework resulted from responses to this question. Both of the pilot participants provided clear and detailed feedback about criteria that they found unclear, unnecessary or that were not present, but that they felt should be included.

6.1.1.3 Did you face any other problems while taking the survey?

PP2 suggested that since all questions of a section 'must' be answered, a clearer notification should be presented to participants in cases where they had omitted to answer certain questions in a section. He further added that the researcher should consider providing the option to allow a participant to leave a question unanswered, as an alternative, should they prefer not to provide an answer to a question.

6.1.1.4 Were you able to save the survey if not completed in one session, and return to it later?

Both participants completed the pilot framework in one sitting, hence could not comment on this question. However, PP2 suggested participants should be encouraged at the outset to complete the framework in its entirety.

6.1.1.5 How long, in minutes, did it take you to complete the pilot framework?

Both pilot participants took longer than anticipated - 1 hour, instead of 30 minutes. However, both indicated that this was due to the meticulous nature of their interaction, as they assimilated, analysed and criticised the pilot framework. Both were satisfied that the suggested 30 minutes was sufficient for completion only.

6.1.1.6 Please mention any ways in which I can improve the framework.

The suggestions given in response to this question were mainly with regard to format and presentation. PP1 suggested that since the landing page was a little verbose and text-heavy, it should be divided into sections with some formatting. He further added that the references provided at the top of the framework to substantiate that certain criteria had emerged from the literature, were unnecessary and should be removed. The final comment was that the font size used when presenting each option was substantially smaller than the body text, thus this should be increased.

PP2 suggested, that the format of the framework could include additional highlighting, so that respondents could see at the outset what sections/topics would be addressed.

6.1.1.7 Changes effected to the Pilot Framework

The changes suggested by the pilot participants in Study 4a were effected by the researcher:

- rewording some of the actual criteria for better clarity,
- moving criteria to other categories where a better fit was identified,
- including explanations of ambiguous terms used within the criteria and, in some instances, providing examples to elaborate the criteria,
- removing similar criteria across and within categories to avoid duplication, and
- incorporating additional criteria that were necessary but that had not been included in the Pilot Study.

The suggestions outlined by PP1 and PP2, summarised in Table 6.3, were worked into the development of the next version of the framework, the Evaluation Framework. The table lists the participants' feedback and suggestions, referring to the section of the Pilot Framework to which it related, and indicates the researcher's response. The resulting version, the Evaluation Framework, as illustrated in Figure 6.2, was the input to Study 4b, the Evaluation Study, which is discussed in Section 6.1.2.

Table 6.3: Changes to Pilot Framework based on pilot participants' reviews, Study 4a

Category Number	Category	Suggestions by participants in Pilot Study	Criterion reference in original Pilot Framework	Criterion reference in Evaluation Framework	Participant Code	Researcher's response
1	Interface Design	Include facility for the student to return to questions left unanswered, incomplete or requiring a change of answer			PP2	All questions compulsory – structured to complete all in the category before moving to next question
		Include facility to skip a question and return later			PP2	All questions compulsory – structured to complete all in the category before moving to next question
		Include facility to delete the given answer and revert question status to unanswered			PP2	Happens automatically when students change an answer, but it cannot be left blank
		Include time keeping for students		10	PP2	Criterion inserted in this section
2	Question Editing	The term — Range of "parameters" is not clear	2.8		PP2	Criterion reworded for better clarity, with example to illustrate
		The term, "question metadata" is not clear	2.10; 4.4		PP1, PP2	Criteria reworded for better clarity, with example to illustrate
		Include facility to send comments to the academic	2.18		PP2	Sending to academic and author is the same, hence the criterion was reworded.
3	Assessment Strategy	It should force this situation, so that tests taken at different times do not become compromised immediately the first students write the on-line test, especially for summative assessment.	3.1; 3.7		PP2	All questions compulsory – structured to complete all in the category before moving to next question
		For consideration: Can a student take the test at different times for different sections?		9	PP2	Criterion included in Assessment Strategy section for Self Assessments

		The preamble does not make it clear which aspects are for formative and which for summative assessments	3.7		PP2	Preamble reworded for more clarity
		The term "branching of questions" is not clear	3.3		PP2	Criterion reworded for better clarity, with example to illustrate
		Include facility to delete the given answer and revert question status to unanswered	3.6		PP2	All questions compulsory – structured to complete all in the category before moving to the next question, thus a criterion cannot be left unanswered
4	Test and Response Analysis	Include facility to obtain the normal statistics that academics require		43 and 44	PP2	2 Criteria included in this section
6	Security	If an answer cannot be changed by a student during an assessment (through an option that is specifically "turned on" for that test/section only and where the student is clearly informed of the situation) then include an "Are you Sure" Message Box to handle the case where a key/mouse button is accidentally pressed.	6.9		PP2	The Framework is designed for usage by academics and the current version is not intended for student usage. If required in the Student version, it would be a design feature to include.
		Re-phrase so that the intent becomes clear on the first reading of the question	6.10		PP2	Criterion reworded for better clarity
8	Ease of Use	Criteria request the same information	8.17		PP2; PP1	Criterion deleted
		Questions have no option to allow the facility to academics but deny it to students – there is no facility to distinguish.		17 to 23	PP2	New criteria created to focus on facilities required by academics, which may be different to student requirements
9	Technical Support Criteria	Does this mean "automatically enroll" or "self-enroll"?	9.12		PP2; PP1	Criterion reworded for clarity
		Unless you mean "relevant to that academic", there is no need, since the academic may not be interested in that feature.	9.15		PP2; PP1	Criterion reworded for clarity

Following this pilot study, adjustments were made to the Pilot Framework. Eleven criteria (Appendix G: Interface Design - Criterion 10; Assessment Strategy – Criterion 9; Test and Response Analysis – Criteria 43 and 44; Ease of Use – Criteria 17 to 23) were added in and seven (Appendix F2: Ease of Use – Criteria 1, 9 to 14) moved/removed. This resulted in the next version of the SEAT Framework, namely, the Evaluation Framework, containing 166 criteria, in ten categories. This Evaluation Framework was to be used in the Evaluation Study, Study 4b in Section 6.1.2. This version of SEAT is provided in Appendix G.

6.1.2 Study 4b – Evaluation Study

Following the refinements to the Pilot Framework of SEAT, as presented in Table 6.3 in Section 6.1.1, the next version of the SEAT Framework, the Evaluation Framework, depicted in an earlier figure, Figure 6.2, was generated and evaluated in Study 4b. This study is summarised in Table 6.4 (similar to a portion of Table 4.8 Table 4.8 in Section 4.5.4). It contributed to answering Research Question 5.

Table 6.4: Summary of Study 4b as outlined in Figure 6.1

Study 4b (April – May 2012) – Evaluation Study	
Respondents	56 Participants from 16 Universities in South Africa
Data Collection	SEAT Instrument (See Appendix G)
Data Analysis	Quantitative – basic statistical analysis Qualitative – content analysis
Purpose	To determine which of the criteria identified in the literature reviewed in Phase 1, are essential for any e-assessment tool

Subsequent to the Pilot Study, the updated Evaluation Framework was then ready to be distributed as part of the Evaluation Study. The purpose of this major and extensive study was to continue the action research process of refining the SEAT Framework as a prelude to the development of an electronic framework for evaluating and adopting e-assessment tools.

Eighty (80) participants were identified for the Evaluation Study. They were selected from the groups of participants who had participated in Studies 1, 2 and 3 (questionnaires and interview sessions) in Phase 1 of this research.

Only a few responses were received by the initial collection date, therefore the researcher made personal follow-up calls to participants to encourage them to take part in this study. Email invitations with the link to the survey: http://www.surveymonkey.com/s.aspx were distributed via the Survey Monkey online survey tool. A total of 53 complete responses and three partial, though usable, responses were received over a three week period. This resulted in a 70% response rate.

The Evaluation Framework, which is provided in Appendix H, presented evaluation criteria grouped into the same categories as in the Pilot Study. For each criterion in each category, the respondent was asked to rate the criterion on a scale of 1 to 7, where 1 = Extremely Important and 7 = Not at all Important. An extract is given in Figure 6.4, with the category title ('Interface Design') at the top and the criteria in the category listed below it.

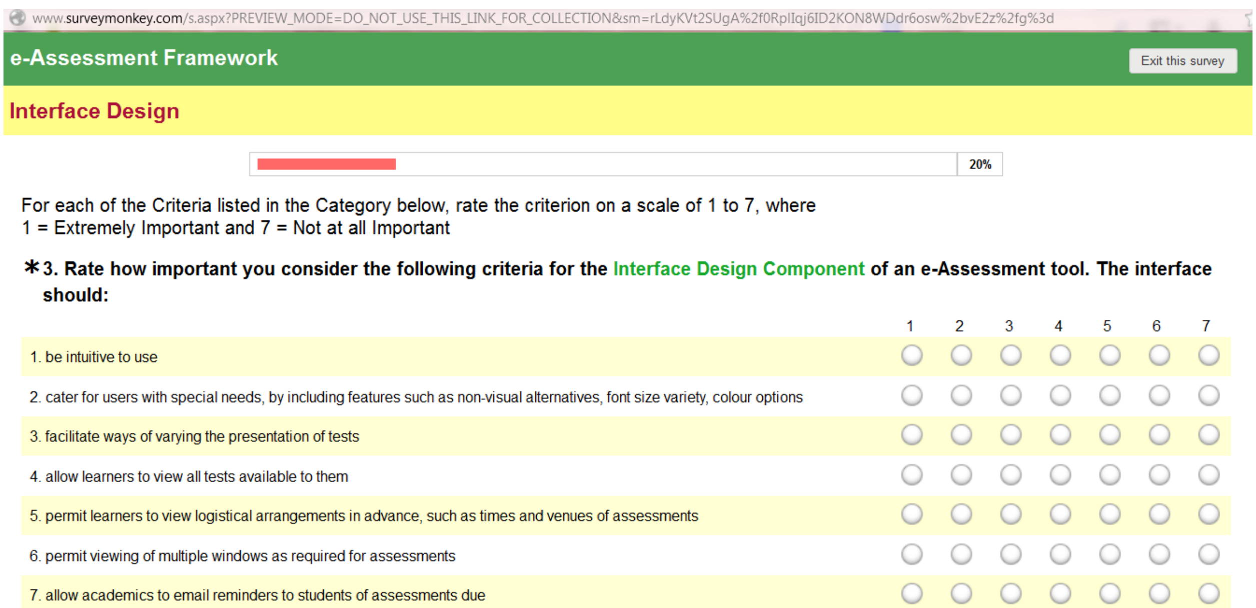


Figure 6.4: Sample of a screen for the Interface Design Criteria in Survey Monkey illustrating how participants had to complete the SEAT Evaluation Framework for the Evaluation Study

The data collected in this study was statistically analysed with the non-parametric sign test, to assist the researcher in identifying the criteria that were essential to include in the framework. The first round of statistical analysis involved doing a mean calculation on the ratings given by participants to each criterion. These mean scores indicated which of the items were least preferred and which were most preferred. The interpretation was that if a mean score was significantly ≤ 3 then the item was relevant. If the mean score was significantly ≥ 6, then the item should be removed.

Thus the following three groupings were created by the researcher for analysis purposes:

- *A*: Where values of 1, 2, 3 indicated that criteria should be retained,
- *B*: Where values of 4 and 5 indicated criteria that the researcher should review and decide whether or not to retain them, and
- *C*: Where values greater than 6 indicated that criteria should be removed.

The results of this statistical analysis are presented in Tables 6.5 to 6.14.

For each of the ten categories listed in Section 6.1.1, the upcoming tables in Section 6.1.2.1 to 6.1.2.10 respectively, present all of the criteria initially in that category. The text that follows each table demonstrates the means and standard deviations of the ratings assigned to each criterion, and explains which criteria were retained following the rating process, and which could be removed. A second important role of the tables is that they provide us with the first comprehensive presentation of the categories and criteria in the SEAT Framework.

It must be noted that all 56 participants completed the evaluation of the first two categories, but thereafter the number of participants decreased marginally. Categories 6, 7, 8, 9 and 10 were completed by 53 of the 56 participants.

6.1.2.1 Category 1: Interface Design

The Interface Design category evaluated aspects such as the intuitiveness, usability and communication features required in an e-assessment tool. In this category all the mean values were significantly ≤ 3. They all fell into Group A, as shown in Table 6.5; hence all the criteria were retained.

Table 6.5: Statistical Analysis of the Interface Design Criteria

The software should …	N	Mean	Std. Deviation
1. be intuitive to use	56	1.43	0.828
2. cater for users with special needs, by including features such as non-visual alternatives, font size variety, colour options	56	2.48	1.236
3. facilitate ways of varying the presentation of tests	56	2.57	1.373
4. allow students to view all tests available to them	56	2.43	1.582
5. permit students to view logistical arrangements in advance, such as times and venues of assessments	56	2.21	1.498
6. permit viewing of multiple windows as required for assessments	56	2.63	1.567
7. allow academics to email reminders to students of assessments due	56	2.21	1.398
8. clearly display marks for each question	56	1.52	0.894
9. clearly display marks for each section	56	1.82	1.081
10. display a clock to keep track of time allocated/remaining	56	2.09	1.676

6.1.2.2 Category 2: Question Editing Criteria

The Question Editing category is extensive and plays a vital role in supporting academics who are generating questions. This category presents features that facilitate the creation of questions within e-assessment tools. Some essential characteristics of this category include allowing the academic to:

- add or create his/her own questions for inclusion in the database,
- view/edit/adapt existing questions in the database,
- import/export questions from other systems,
- insert metadata into questions for logical filing and easier extraction, and
- pilot test the assessment prior to students taking the assessment.

All the mean values were significantly ≤ 3, as illustrated in Table 6.6. They all fell into Group A hence all the criteria were kept.

Table 6.6: Statistical Analysis of the Question Editing Criteria

The software should …	N	Mean	Std. Deviation
1. allow the academic to create the test electronically	56	1.41	1.203
2. incorporate procedures that update records immediately, and not at the end of the session, when questions are edited/authored	56	1.89	1.231
3. permit the academic to author original questions to add to the question bank	56	1.48	1.175
4. allow the academic to view existing questions in the question bank	56	1.43	1.158
5. allow the academic to adapt existing questions in the question bank	56	1.48	1.175
6. support importing of questions in non-proprietary, interoperable format to the question bank	56	1.89	1.303
7. support exporting of questions in non-proprietary, interoperable format from the question bank	56	2.11	1.41
8. permit a range of parameters/options to be specified in questions (for example, four or five options per question)	56	1.91	1.468
9. support feedback creation for each question	56	1.95	1.367
10. allow the incorporation of question metadata (for example, categories, keywords, learning objectives, and levels of difficulty)	56	2.34	1.431
11. facilitate offline question creation	56	2.11	1.448
12. grant academics previews of assessments created both offline and online	56	1.98	1.368
13. incorporate an automatic grammar check facility	56	2.64	1.656
14. incorporate a spell checker	56	2.5	1.695
15. assign a global unique identifier to all questions created or revised in the question bank	56	2.3	1.374
16. flag questions which students have not answered in an assessment so that these can be deleted or amended by the academic	56	2.38	1.383

17. allow the academic to approve or reject all created questions before adding to/rejecting from the question bank	56	2.3	1.464
18. allow the academic to add comments to a question created before adding to/rejecting from the question bank	56	2.68	1.281
19. direct comments regarding questions submitted to the question bank directly to the author of the question	56	2.68	1.177
20. allow academics to create a marking scheme for an assessment	56	1.89	1.216
21. allow academics to combine questions from different test banks into a single test	56	1.68	1.281
22. allow academics to pilot tests prior to the assessment going live	56	1.55	1.22
23. support printing of tests	56	1.98	1.328
24. display time taken by the individual student for each question	56	2.84	1.581
25. display time taken by the average student for each question	56	2.82	1.642
26. facilitate allocation of marks to questions	56	1.52	1.062

6.1.2.3 Category 3: Assessment Strategy Criteria

The Assessment Strategy criteria relate to aspects that facilitate easy compilation of tests, especially multiple versions of the same assessment, to reduce the time and effort associated with this type of compilation. In this category all the mean values were significantly ≤ 3, as indicated in Table 6.7. They fell into Group A hence all the criteria were retained.

Table 6.7: Statistical Analysis of the Assessment Strategy Criteria

The software should …	N	Mean	Std. Deviation
1. support random generation of questions from the test bank in multiple versions of the same assessment	55	1.93	1.317
2. support randomisation of sections in multiple versions of the same assessment	55	2.04	1.319
3. incorporate branching of questions, depending on user's response (for example, if user selects option (a) questions 5 to 10 are displayed, else questions 11 to 15 are displayed)	55	2.42	1.462
4. display feedback as/if required	55	1.8	1.325
5. display results as/if required	55	1.8	1.223
6. specify how many attempts a student is permitted to make on a question	55	1.95	1.353
7. permit students to sit a test as many times as they like (in the case of self-assessments)	55	1.85	1.38
8. allow students access to a test without authentication (in the case of self-assessments)	55	2.95	2.05
9. permit a student to take the test at different times for different sections (in the case of self-assessments). (for example, complete section A today, section B tomorrow and eventually complete paper when he/she has the time)	55	2.31	1.359

6.1.2.4 Category 4: Test and Response Analysis

Criteria related to the statistical analysis of assessment responses, are presented in this category, as shown in Table 6.8. Criteria 11, 13, 15, 41 and 42, which are highlighted, returned mean values significantly > 3 but < 5. They therefore fell into Group B. These criteria were therefore reviewed by the researcher and a second round of statistical analysis was conducted by the researcher to aid the decision-making process. Upon further review, the researcher deemed these to be non-essential criteria but did not remove them from the framework at that stage, in anticipation of remaining substudies, which would include experts in the field as participants in the evaluation process.

Table 6.8: Statistical Analysis of Test and Response Analysis Criteria

The software should ...	N	Mean	Std. Deviation
1. permit student access to be revoked, if necessary, while preserving performance data	54	2.39	1.28
2. allow groups to be set up	54	1.98	1.296
3. allow students to be added to a group	54	2.06	1.338
4. permit questions to be viewed by metadata fields (for example, categories, keywords, learning objectives, and levels of difficulty)	54	2.15	1.204
5. allow students access to previous assessment results	54	2.02	1.296
6. allow students access to previous assessment responses	54	2.04	1.213
7. allow students access to markers' comments on prior assessments (in cases where a human assessor reviewed the completed test)	54	2.02	1.124
8. present results immediately to students, when appropriate	54	1.44	0.883
9. allow results to be accessed after a specific date, as required	54	1.96	1.303
10. support the combination of marks with marks from other assessments	54	2.17	1.539
11. allow students to compare the results they obtained with other students' results	54	3.76	2.074
12. allow students to compare marks with group averages	54	2.67	1.625
13. permit students access to answers of other students in an assessment to verify their results	54	4.04	2.128
14. provide students with the option/facility to print out assessment responses	54	3.00	1.614
15. allow students to comment on individual questions	54	3.13	1.614
16. allow students to comment on tests overall	54	2.96	1.479
17. distribute assessors' comments through the system to students	54	2.13	1.166
18. distribute assessors' comments through email to students	54	2.56	1.436
19. prompt for a reason when students make late submissions of self-assessments	54	2.98	1.608
20. provide a warning when students make late submissions of self-assessments	54	2.61	1.642
21. display/present feedback on tests and questions from students to academics	54	2.52	1.463
22. where appropriate, forward feedback related to particular questions directly to the student	54	2.28	1.406
23. permit markers to add notes about students, relating to their responses	54	2.31	1.096
24. email human assessors automatically if marking deadline is not met	54	2.48	1.489
25. present mean (average) score statistical analysis per assessment	54	1.91	1.033

26. present discrimination index statistical analysis per assessment	54	2.15	1.204
27. present facility index statistical analysis per assessment	54	2.33	1.303
28. present highest score statistical analysis per assessment	54	2.24	1.212
29. present lowest score statistical analysis per assessment	54	2.26	1.247
30. present frequency distribution statistical analysis per assessment	54	2.07	1.195
31. incorporate an automated 'cheating spotter' facility	54	2.31	1.464
32. support the ordering of the results tables in various ways (for example, by marks, student numbers, names)	54	2.06	1.497
33. display marks as percentages	54	1.72	1.123
34. present to the academic all attempts at a question	54	2.67	1.66
35. permit the academic to view individual responses to questions	54	1.96	1.288
36. allow the student to view the whole test, as he/she had completed it	54	1.67	1.099
37. support the calculation of grades over a series of tests	54	2.13	1.505
38. display a comparison of mark data of different groups	54	2.41	1.596
39. display a comparison of the performance in different subtopics/sections	54	2.3	1.298
40. permit mark data to be viewed without having access to names of students	54	2.26	1.403
41. support correlation of assessment data with age data	54	3.65	1.803
42. support correlation of assessment data with gender data	54	3.72	1.816
43. flag questions which were poorly answered	54	1.78	1.003
44. flag questions which were well answered	54	1.91	1.086

6.1.2.5 Category 5: Test Bank Design Criteria

In this small but important category the focus was on the databank. The mean values of the criteria in this category were significantly ≤ 3, as shown in Table 6.9. They fell into Group A hence both the criteria remained.

Table 6.9: Statistical Analysis of Test Bank Design Criteria

The software should …	N	Mean	Std. Deviation
1. draw random questions from a question bank, as required	54	1.7	1.268
2. permit students to 'create queries regarding questions' as their responses	54	2.93	1.564

6.1.2.6 Category 6: Security Criteria

The essential security features in an e-assessment tool were presented in the Security Criteria category. This category described elements that would ensure that implementing an electronic assessment is as secure as administering a traditional written assessment. In this category all the mean values were significantly ≤ 3, as shown in Table 6.10. They fell into Group A hence all the criteria remained.

Table 6.10: Statistical Analysis of Security Criteria

The software should …	N	Mean	Std. Deviation
1. ensure that tests are accessible only to users who have explicit authorisation, granted by access administrators	53	1.47	1.049
2. encrypt all data communicated via the network	53	1.85	1.277
3. ensure that mark data held on the server can be accessed by authorized persons only	53	1.47	1.012
4. be robust and prevent 'crashes' on the server or browser	53	1.34	.999
5. log where students sat	53	2.96	1.593
6. log the IP address where each student sat	53	2.68	1.516
7. log which questions were marked by which human assessor	53	2.28	1.392
8. log when the marker marked the question	53	2.81	1.653
9. prevent alterations of answers to questions already completed (in cases where second opportunities are not permitted)	53	1.87	1.415
10. require permission of the question author before any question can be modified or deleted from a test	53	2.60	1.680
11. prevent students from amending a test once taken	53	1.74	1.288
12. prevent students from deleting a test once taken	53	1.83	1.438
13. automatically allocate a global unique identifier to tests	53	1.89	1.340
14. provide ability to view entire tests for verification without the ability to change them	53	2.11	1.296
15. restrict tests to particular IP addresses and domains	53	2.02	1.525
16. allow academics to enter details of students who cheat	53	2.47	1.488
17. permit academics to modify results after communication with a student regarding the reason for the change	53	1.94	1.082
18. log and motivate modifications to original marks	53	1.75	1.142

6.1.2.7 Category 7: Compatibility Criteria

The Compatibility Criteria outlined in this category addressed the aspects that facilitate the integration of an e-assessment system with existing institutional systems. In this category all the mean values were significantly ≤ 3, as shown in Table 6.11. They fell into Group A hence all the criteria remained.

Table 6.11: Statistical Analysis of Compatibility Criteria

The software should …	N	Mean	Std. Deviation
1. be accessible from a standard, platform-independent web-browser, without additional plugins	53	1.47	1.219
2. be downgradable for users with previous versions of browsers	53	2.00	1.316
3. be customisable to provide a uniform interface with the rest of the institution's intranet or virtual learning environment	53	1.91	1.377
4. link seamlessly with other institutional systems, so that users can use their existing username and passwords	53	2.00	1.468
5. permit results to be exported to spreadsheets or statistical analysis software	53	1.47	1.295
6. support entry of marks, dates of submission and other details of non-computer-aided assessments	53	2.08	1.627
7. link seamlessly with other institutional systems so users can share student details	53	2.23	1.540
8. link seamlessly with other institutional systems so users can export marks directly	53	2.13	1.545

6.1.2.8 Category 8: Ease of Use Criteria

The Ease of Use Criteria presented elements that aided in making the e-assessment system user-friendly, and thus facilitated the ease with which novice users could adopt the system. In this category, Criterion 12 had a mean value significantly > 3 but < 5, as shown in Table 6.12. It therefore fell into Group B. This criterion was therefore reviewed by the researcher. Upon further review, the researcher deemed this to be a non-essential criterion but did not remove it from the framework, in anticipation of remaining substudies, which would involve experts in the field as participants in the evaluation process.

Table 6.12: Statistical Analysis of Ease of Use Criteria

The software should …	N	Mean	Std. Deviation
1. require little time to capture data related to student profiles and assessments	53	1.83	1.221
2. require a short time period to set up an assessment online	53	1.79	1.098
3. require little/no training on how to use the tool	53	1.87	1.075
4. provide simple and fast login procedures	53	1.64	1.002
5. include an intelligent help system – dependent on the user role and current activity	53	2.15	1.063
6. incorporate speech synthesis for special-needs users	53	2.70	1.539
7. be intuitive to use – users should not require any special programming language skills to adopt the tool	53	1.45	.774
8. make it easy to include multimedia elements in test items	53	1.66	1.143
9. allow academics access to details of room numbers/venues of an assessment	53	2.75	1.426
10. allow academics access to details of times of an assessment	53	2.49	1.476
11. permit all students in a group to be removed from the system simultaneously	53	2.38	1.417
12. allow academics access to details of invigilators for an assessment	53	3.06	1.549
13. allow access to details of students sitting a test at a particular time	53	2.53	1.324
14. permit students to return to the point at which they had exited an incomplete self-assessment test	53	1.83	1.087
15. make it easy, where necessary, to enter foreign characters and symbols	53	1.83	.975
16. automatically distribute electronic certificates of test submission to students	53	2.57	1.487
17. allow students access to details of room numbers/venues of an assessment	53	2.32	1.425
18. allow students access to details of times of an assessment	53	2.08	1.385
19. simplify the task of adding user access	53	1.87	.941
20. simplify the task of removing user access	53	1.89	1.013
21. simplify the task of editing user access	53	1.87	.962
22. allow students and other users to be enrolled on the system by an administrator	53	1.79	1.116
23. allow students and other users to be removed from the system	53	1.79	1.081

6.1.2.9 Category 9: Technical Support Criteria

The technical aspects of an e-assessment system were outlined in the Technical Support Criteria category. In this category all mean values were significantly ≤ 3, as shown in Table 6.13. Thus they fell into group A. Hence all the criteria remained.

Table 6.13: Statistical Analysis of Technical Support Criteria

The software should …	N	Mean	Std. Deviation
1. incorporate a resilient network	53	1.72	1.199
2. if not web-based, include software that is easy to install, requiring little effort and time	53	1.58	1.082
3. run on Windows and UNIX servers	53	2.02	1.500
4. include installation software that is easily available	53	1.57	1.029
5. allow new functionality to be incorporated without reinstalling the system	53	1.62	.985
6. support large numbers of concurrent users logged in simultaneously	53	1.23	.640
7. support multi-format data storage – Oracle/Access or ODBC (Open DataBase Connectivity) format	53	1.92	1.342
8. facilitate the use of existing database systems	53	2.04	1.427
9. grant academics access to details of all test purchases relevant to that academic, where tests are purchased from the supplier of the assessment software	53	2.38	1.431
10. facilitate the transfer of sales and purchaser details to separate e-commerce systems	53	2.98	1.550

6.1.2.10 Category 10: Question Types

The large variety of MCQ formats supported by e-assessment systems were presented in the Question Types category. In this category all mean values were significantly ≤ 3, as shown in Table 6.14. Thus they fell into group A. Hence all the criteria remained.

Table 6.14: Statistical Analysis of Question Types Criteria

The software should support …	N	Mean	Std. Deviation
1. Multiple choice: single response	53	1.28	.841
2. Multiple choice: multiple response	53	1.64	1.287
3. True/False	53	2.19	1.755
4. True/False with explanation	53	2.04	1.344
5. Fill-in-the-Blanks/Completion	53	2.09	1.431
6. Simulation	53	2.08	1.371
7. Matching Items	53	2.17	1.464
8. Extended-Matching Items	53	2.09	1.275
9. Selection/Drop Down Lists	53	1.94	1.247
10. Ranking	53	2.21	1.561
11. Diagram/Video Clips	53	1.89	1.103
12. Drag-and-Drop	53	2.42	1.460
13. Reordering/Rearrangement/Sequencing	53	2.36	1.469
14. Categorising	53	2.26	1.361
15. Hotspots	53	2.62	1.584
16. Text Input (short answer – which would be marked manually)	53	1.94	1.200

6.1.2.11 Summary of first-round analysis using groupings A, B and C

In the ten categories presented in Study 4b, there were a total of 166 criteria (see Table 7.2) for participants to review.

Following this first-round statistical analysis, all the criteria that were deemed non-essential, based on the mean values calculated - namely 11, 13, 15, 41, 42 (in Category 4: Test and Response Analysis) and 12 (in Category 8: Ease of Use) – were not removed, as the researcher was still to conduct two more substudies (Study 4c and 4d), which would eventually assist in determining if these 'possible deletion' criteria should actually be removed from the framework.

To further validate the framework and determine whether other criteria should be eliminated, or considered for removal, another set of statistical analysis, namely, the Non-parametric Mann-Whitney Test, was conducted, which is explained in Section 6.1.2.12.

6.1.2.12 Second round analysis using further groupings

Further statistical analyses were done on the same dataset, but using two sets of groupings, namely, considering respondents who were Computing users versus those were Non-Computing and, similarly, those who were Academic users versus those who were Non-Academic. In addition to the A, B and C groupings created in the introductory part of Section 6.1.2, further groupings were defined by the researcher for analysis purposes:

- D: Analysis of ratings of Computing versus Non-Computing users, and
- E: Analysis of ratings of Academic versus Non-Academic users.

In both cases, analysis was carried out using the Non-parametric Mann-Whitney Test to ascertain whether a significant difference existed between the results from Computing and Non-Computing users (Grouping D) and, similarly, between the results from Academic and Non-Academic users (Grouping E). If a certain criterion returned a significant difference between the ratings of the two sets of users, and Non-Computing users or Non-Academic users found it more important, then that criterion was considered for removal. But if Computing or Academic users found it more important, then it remained. Where a significant difference occurred, both the mean values were used to determine whether the criterion should be retained or removed. For example, if it was a Grouping D analysis, the mean values of the Computing rating and the Non-Computing rating would be investigated.

Participants were not required to re-take the Evaluation Framework survey on Survey Monkey. The same results that were generated from their initial completion of the Evaluation Study, as illustrated in Figure 6.5, were used to conduct this analysis.

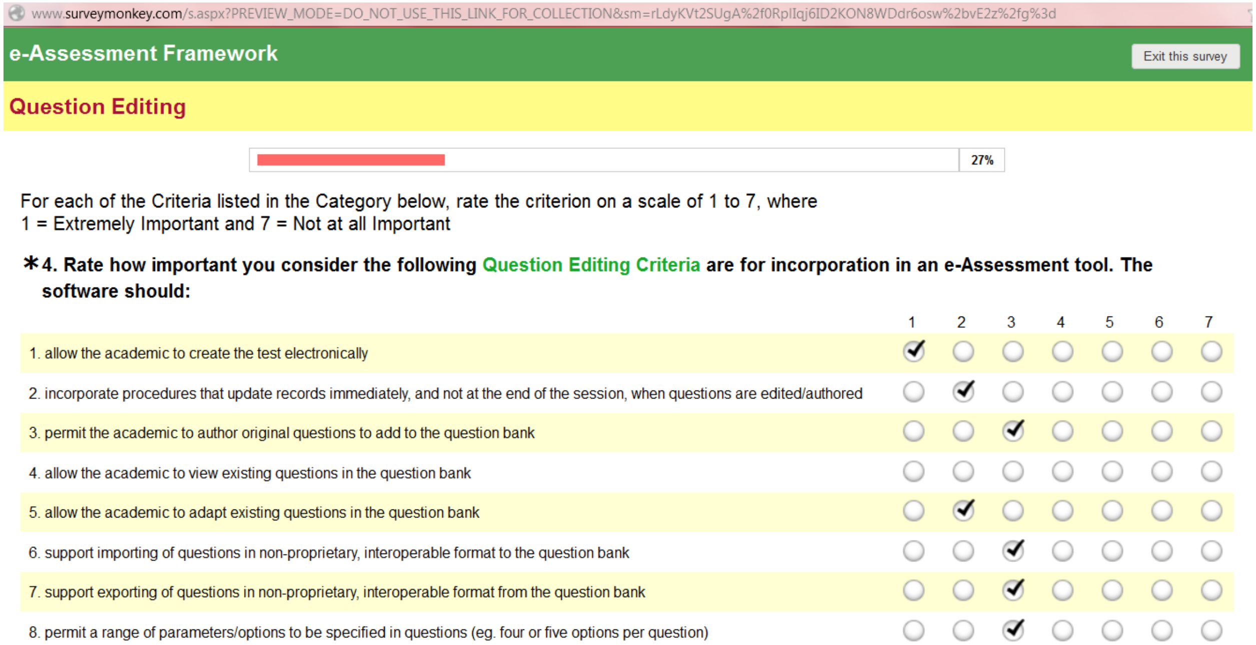


Figure 6.5: Sample of the Question Editing Criteria screen in Survey Monkey illustrating how participants had to complete the SEAT Evaluation Framework for the Evaluation Study

For a Mann-Whitney test to be significant, mean values must be significantly greater than 3 for the criterion to fall into the 'important to include' grouping (Group A in introductory part of Section 6.1.2). Where this was not the case, the researcher reviewed them to take an informed decision on whether or not they should be included (Group B in introductory part of Section 6.1.2). Furthermore, if there was a significant difference between the ratings of the two sets of users in the grouping under consideration, the researcher considered the directionality of the difference and thence decided whether that criterion should remain in subsequent versions of the framework.

However, in both sets of groupings, D and E, no criteria had significant p values following a Mann Whitney test, that is, no mean scores were significantly different. Therefore, no criteria were identified for removal under Mann-Whitney. The researcher then decided, after inspection, that certain criteria where there were notable (though not significant) differences between Computing and non-Computing or between Academic and non-Academic, should also be considered as candidates for elimination. These cases are shown in Table 6.15 and Table 6.16.

Table 6.15: Statistical Analysis of Test and Response Analysis Criteria

	Computing			Non-Computing			
The software should …	N	Mean	Std. Deviation	N	Mean	Std. Deviation	Mann Whitney p-value
13. permit learners access to answers of other learners in an assessment to verify their results	36	4.14	2.058	18	3.83	2.307	.616
15. allow learners to comment on individual questions	36	3.00	1.531	18	3.39	1.787	.479
41. support correlation of assessment data with age data	36	3.47	1.699	18	4.00	2.000	.283
42. support correlation of assessment data with gender data	36	3.47	1.647	18	4.22	2.074	.149

Table 6.16: Statistical Analysis of Ease of Use Criteria

	Academic			Non-Academic			
The software should …	N	Mean	Std. Deviation	N	Mean	Std. Deviation	Mann Whitney p-value
12. allow academics access to details of invigilators for an assessment	43	3.02	1.551	10	3.20	1.619	.789

This analysis identified Criteria 13, 15, 41, 42 (in Category 4: Test and Response Analysis) as given in Table 6.15 and Criterion 12 (in Category 8: Ease of Use) as given in Table 6.16, as candidates for elimination. Criterion 13 was rated higher by Computing users than by non-Computing, so it remained. Upon reviewing criteria 15, 41 and 42, the researcher marked them as 'consider for removal', since they could be regarded as non-essential criteria and were rated higher by non-Computing users. Criterion 12 was also marked by the researcher, after review, as 'consider for removal', since it was rated higher by non-Academic users and can be regarded as non-essential.

These four candidates for removal had already been identified as non-essential criteria in the initial analysis (Sections 6.1.2.4 and 6.1.2.8), so this analysis played a confirmatory role. However, it must be reiterated that these non-essential criteria were not removed at this stage, as the researcher was still to conduct two more substudies (Study 4c and 4d), which would eventually assist in determining if these 'possible deletion' criteria should actually be removed from the framework.

To summarise, the initial round of statistical analysis conducted in Study 4b on the Evaluation Framework that was distributed to respondents had two aims, namely:

- to determine whether an item was relevant for inclusion or not in the framework. This was determined by the mean values of criteria falling into one of three categories:
    - *Group A*: Mean≤3: Item is relevant
    - *Group B*: Mean≥6: Item is irrelevant, hence remove
    - *Group C*: Mean>3 but <6: Researcher's decision on whether to include or remove
- to determine if a difference existed between the perceptions of the two subcategories of respondents regarding an item's relevance:
    - *Group D*: Computing vs Non-Computing
    - *Group E*: Academic vs Non-Academic

This concludes discussion of the results from the statistical analysis conducted in Study 4b.

The qualitative comments made by the respondents were taken into consideration and are briefly discussed in Section 6.1.2.13.

6.1.2.13 Qualitative Comments

When asked, in open-ended questions, for general comments or concerns, only seven of the 56 respondents completed the open-ended sections.

A number of respondents indicated that the tool was comprehensive and useful. R2 stated that "All statements are obvious rules for e-assessment". R4 was much more enthusiastic, saying, "These are fantastic criteria you have selected which I think will enable developers of open-source systems to customise and enhance their tools". This was supported by R14 who said, "There was not a single item that I would not want the option of including in an assessment tool system". These statements indicate their perceptions that the researcher had identified crucial items for inclusion in an e-assessment tool.

In R28's experience, "the survey made us also re-think our online assessment and its alignment with our lecturing" while R32 found it "an interesting survey".

Only two respondents (R24 and R46) "found the questionnaire to be too long".

6.1.2.14 Evolution of the SEAT Framework

The following changes, summarised in Table 6.17, were made to the SEAT Framework, based on the statistical and manual analysis, as well as the qualitative comments in Study 4b.

Table 6.17: Changes made to SEAT Framework after the main evaluation in Study 4b

Category number	Category Name	Criterion reference in original Evaluation Framework	Researcher's response
1	Interface Design	8, 9, 10	Reworded for better clarity
		11, 12, 13, 14	Created and added items that respondents deemed important for an e-assessment system
2	Question Editing	12	Split into two individual items, one for offline and the other for online
		18, 23, 26	Reworded for better clarity
		27 to 33	Created and added items that respondents deemed important for an e-assessment system
3	Assessment Strategy	11 to 15	Created and added items that respondents deemed important for an e-assessment system
4	Test and Response Analysis/Reports	14, 45-48	Created and added items that respondents deemed important for an e-assessment system
5	Test Bank	3	Created and added an item that respondents deemed important for an e-assessment system
		4	Originally Criterion 15 in the Question Editing Category; moved to Test Bank Category for improved contextualisation
6	Security	18, 21-23	Created and added items that respondents deemed important to include in an e-assessment system
		19	Split into two individual items, one to log modifications and the other for recording motivations for these changes
7	Compatibility	9	Created and added an item that respondents deemed important to include in an e-assessment system
8	Ease of Use	24, 25	Created and added items that respondents deemed important to include in an e-assessment system
10	Question Types	16	Created and added an item that respondents deemed important to include in an e-assessment system

Across all ten categories, based on the statistical analyses undertaken (non-parametric sign test and Mann-Whitney test), six criteria were noted for possible removal – namely 11, 13, 15, 41 and 42 (in Category 4: Test and Response Analysis) and 12 (in Category 8: Ease of Use). The changes (32 additional criteria) specified in Table 6.17, were used to evolve the SEAT Framework into the next version (with 198 criteria), namely the Proof of Concept Framework (see Appendix H1). This Proof of Concept Framework would be reviewed by experts in the field of e-assessment and MCQs. This was carried out in Study 4c, the Proof of Concept Study, which is addressed in Section 6.1.3.

6.1.3 Study 4c – Proof of Concept Study

After the pilot evaluation in Study 4a and the main evaluation in Study 4b, the improved version of SEAT reached the stage of an operational prototype, demonstrating a proof of concept. At this stage, SEAT had not reached an electronic stage of operation; it was still a *Survey Monkey* version that was used. Oates (2010) describes a Proof of Concept (PoC) pointing out that not all researchers actually undertake evaluations of a system or artefact they have designed. Instead, they might just show a PoC by generating a prototype that functions, and that behaves in a required way under certain conditions.

In the case of this action research process, however, there was both an evaluation and a functioning PoC. Study 4b (Section 6.1.2) was an extensive evaluation study with many participants reviewing and evaluating the SEAT Framework in an effort to improve it. This improved and refined version of SEAT became the Proof of Concept Framework. In the rigorous investigation and development of SEAT, the PoC stage was extended to include a small-scale qualitative evaluation of SEAT by taking the PoC version to three experts in the fields of e-assessment and MCQs. This became Study 4c, the Proof of Concept Study, as outlined in Figure 6.2 and Table 6.18 (which was initially shown as part of Table 4.8 in Section 4.5.4). It contributed to answering Research Question 5. The three experts identified, brought in three different dimensions to reviewing the framework, as Participant One (PoC1) was an e-learning manager, (recommended by the researcher’s supervisor) PoC2 was an academic leader who was responsible for decisions taken regarding the adoption of e-assessment tools in his school (a participant in Study 1 whose responses distinguished PoC2 as being highly knowledgeable in the field of both e-assessment and MCQs), and PoC3 was an academic, who had specialised in MCQs for more than five years (a participant in Study 1 whose responses indicated great expertise in MCQs).

Table 6.18: Summary of Study 4c as outlined in Figure 6.1

Study 4c (May 2012) – Proof of Concept Study	
Respondents	3 expert users (UCT, UNISA and WITS)
Data Collection	SEAT Instrument (See Appendix H1) and Interviews (See Appendix H2)
Data Analysis	Qualitative - content and discourse analysis
Purpose	To gain insight into the criteria regarded as 'essential' for inclusion in the Framework

These three 'experts in the field' were invited by the researcher to critically evaluate and suggest improvements to the framework, prior to applying the next version, the Application Framework, to an existing e-assessment system in Study 4d.

The Proof of Concept (PoC) Framework (with 198 criteria) was hosted once again on Survey Monkey (see Appendix H1). The link to the PoC Framework was then emailed to the three participants, who were given a brief background to the study and requested to evaluate the framework. This was then followed up with personal telephonic interviews with each participant to clarify comments made in their responses. The interview schedule with the three PoC participants can be viewed in Appendix H2. These qualitative responses provided useful verbal data and spontaneous comments that enriched this study.

6.1.3.1 Evolution of the SEAT Framework

Changes were made to improve the SEAT Proof of Concept Framework, based on suggestions made by the participants in this study, Study 4c. This section discusses the most far-reaching refinements and then lists all the changes in Table 6.22.

The most important overarching comment/suggestion was made by PoC3, who outlined that each of the categories identified by the researcher could be further grouped into Functional and Non-Functional criteria. Thus the existing criteria, resulting from Table 6.17, were restructured and allocated to Functional or Non-Functional sections. There were initially ten categories of criteria. Following the suggestion made by PoC3, eleven categories were created, and each category further classified into Functional and Non-Functional Criteria, as presented in Table 6.19. This change resulted in a fundamental structural and semantic improvement to SEAT.

Table 6.19: Recategorisation of criteria in the SEAT Framework

	Functional Criteria		Non-Functional Criteria
1	Question Types	1	Interface Design
2	Assessment Strategy	2	Technical Support
3	Test and Response Analysis	3	Security
4	Test Bank	4	Compatibility
5	Question Editing	5	Ease of Use
		6	Robustness (added in at the suggestion of PoC3)

PoC2 suggested that the rating scale used for measuring the criteria should be adapted. The scale used in the first three versions of the framework is given in Table 6.20.

Table 6.20: Initial rating scale used in the SEAT Framework (Pilot, Evaluation and Proof of Concept Frameworks)

1	2	3	4	5	6	7
1 = Extremely Important						
7 = Not at all Important						

Based on the suggestion made by PoC2, the adapted rating scale was modified to a more qualitative type of ranking, as presented in Table 6.21. Participants were required to evaluate how effectively each criterion serves the tool being rated, on a scale from ‘Very Effectively’ to ‘Not at all’.

Table 6.21: Adapted rating for the SEAT Framework after Study 4c

Rating	Very Effectively	Satisfactorily	To a limited extent	Unsatisfactorily	Not at all	Not applicable (N/A)
Numeric Weight	5	4	3	2	1	0

Since the application of the framework requires participants to indicate which features are present in the e-assessment tool they are investigating, a N/A option was also incorporated for flexibility in cases where a feature was not applicable to the tool being investigated.

Further changes were made to the SEAT Framework, based on the qualitative comments and suggestions made by the participants in Study 4c. These refinements are summarised in Table 6.22.

Table 6.22: Changes made to SEAT Proof of Concept Framework after Study 4c

Category number	Category Name	Criterion reference in original Proof of Concept Framework	Participant code	Researcher's response
1	Interface Design	13,14	PoC1, PoC2	Reworded for better clarity
2	Question Editing	2,24	PoC3	Reworded for better clarity
		24,25	PoC2	Combined as they essentially requested the same information
		28	PoC2	Deleted - participant feedback was not necessary
3	Assessment Strategy	2	PoC1, PoC3	Deleted - Criterion 1 essentially requested the same information
		8	PoC1, PoC2, PoC3	Deleted - student authentication was compulsory and necessary
		11,12	PoC2	Deleted - criteria were not based on sound pedagogical principles
4	Test and Response Analysis/Reports	12,13, 32 to 35	PoC1, PoC2, PoC3	Reworded for better clarity
		34, 36	PoC2	Included as a recommendation from PoC2
		9, 14 - 23, 37	PoC1, PoC2, PoC3	Deleted - not essential features required for an electronic tool
		41, 42	PoC2, PoC3	Deleted - these statistical analyses are not required for an e-assessment tool
5	Test Bank	2,3	PoC2	Initially reworded for better clarity, but eventually deleted - students should not have freedom to criticise questions in an assessment
		4,5	PoC2	Included as a recommendation from PoC2
6	Security	3,4	PoC2,3	Reworded for better clarity. Criterion 4 was subsequently deleted - no e-assessment tool should ever 'crash', testing must be completed upfront for all possible errors and scenarios prior to selling the item
		6	PoC3	Reworded for better clarity
		5,6	PoC1, PoC3	Criteria 5 and 6 were combined into one item, as they essentially requested the same information
7	Compatibility	8	PoC1	Reworded for better clarity
		6	PoC2, PoC3	Deleted - any e-assessment tool should support only its own electronic assessments. They need not include features to support non-electronic assessments. This should rather be handled by the Learning Management System or Student Record System at the institution
8	Ease of Use	20	PoC2	Reworded for better clarity
		9, 12	PoC3	Deleted - venue details of assessments need not be communicated by an e-assessment system
9	Technical Support	3	PoC2	Reworded for better clarity
10	Question Types	12	PoC2	Split into two question types as recommended by PoC2
11	Robustness		PoC3	New category created, as suggested by PoC3
		1 to 5	PoC3	Included, adapted from the robustness criterion deleted previously
General change in Study 4c: Grouping of criteria into Functional and Non-Functional Groupings				

As described in Table 6.22, ten new criteria were added to the SEAT Framework after the Proof of Concept Study, while 28 criteria were deleted. Since participants in Study 4c were leading experts in the field of e-assessment and MCQs, they were able to provide confirmation of deletion of the non-essential criteria identified during the statistical analyses conducted in Study 4b.

6.1.3.2 Overall comments on the SEAT Framework

PoC1 commented on the applicability of the framework indicating that "… in an academic context one might theoretically be able to select the right instrument for the specific need. A tool like this would be awesome. Reality is that the institution will end up having one tool and you must use it for whatever you want it. You will be expected to find work-arounds wherever the current tool does not suffice, so in general although someone might be using a tool like this they might not have the luxury to act on its recommendations ... sometimes many of the ideas/features offered by these assessment engines are high and important when selections are done but seldom used in practice."

However, he further stated that "it is a wonderful idea and might be excellent to guide an institution in decision making but it matters who we are, at what time we are making the decision and, oh yes, I forgot we already have a tool in our current LMS which you just have to use anyway ..."

PoC2 indicated that with a little more refinement ... "there is value" in the framework.

PoC1 was positive about the benefits the framework would bring by "… benefiting most stakeholders, but as indicated above you need to consider the practical aspects."

PoC3 suggested that there were "some shortcomings/gaps", which were discussed in the follow-up telephonic interview with the respondent. These structural suggestions are outlined in Table 6.17 and were worked into the next version of the framework.

Based on the above comments and suggestions summarised in Table 6.22, the SEAT Framework was modified prior to Study 4d, the Application Study (Section 6.1.4). The Proof of Concept version evolved to include the new category of Robustness, thus consisting of eleven categories and 180 criteria (see Table 7.2). The Proof of Concept Framework evolved into the Application Framework.

6.1.4 Study 4d – Application Study

Following the Proof of Concept Study (Section 6.1.3), seven respondents were selected to participate in the Application Study. The purpose of the Application Study, as outlined in Table 6.23, similar to part of Table 4.8, was to allow respondents to apply the Application Framework to an existing e-assessment system. Whereas Studies 4a, 4b and 4c had served to examine the categories and criteria within SEAT, Study 4d investigated SEAT in operation. This was to serve as the last application of the Framework via Survey Monkey, prior to replicating it electronically as e-SEAT (see Appendix J1). The study contributed partially to answering Research Question 6.

Table 6.23: Summary of Study 4d as outlined in Figure 6.1

Study 4d (July 2012) – Application Study	
Respondents	7 expert users (UFS, UP, CPUT, UJ and NWU)
Data Collection	SEAT Instrument (See Appendix I), Questionnaire (See Appendix J2) and Follow-up Interviews, where required
Data Analysis	Qualitative – content and discourse analysis
Purpose	To apply the instrument developed (SEAT – Selecting and Evaluating an e-Assessment Tool) to an existing/adopted e-assessment system

6.1.4.1 Participants selected for Study 4d

The seven respondents included five academics who had used e-assessment extensively over a period of five or more years; one was an e-assessment expert as well as an e-consultant; and another, a leading academic and a researcher in e-assessment. These respondents were carefully selected by the researcher due to their expertise in e-assessment, and their ability to provide constructive criticism on the final framework as they applied it to evaluate an existing e-assessment/MCQ tool.

6.1.4.2. Discussion of findings

Four participants applied SEAT to the e-assessment tool embedded in their respective university's Learning Management Systems (LMS).

Most of the participants (four) implemented e-assessment only for undergraduate students. Two had sufficient confidence in e-assessment to use it for postgraduate level modules, whereas the remaining participant used it at both levels. Medium to large classes (31 to 50 and 51 to 100 students) were the realms where their greatest use of e-assessment occurred (four participants in each category respectively). It is interesting to note that despite the trend of adopting e-assessment for large classes, users in this study also adopted it for smaller classes (<30).

Participants were required to answer a set of qualitative questions after their interaction with the SEAT Framework (Appendix J2) and their resulting comments are listed below. Participants A5, A6 and A7 made no textual comments and were not available for follow-up interviews.

Application of the SEAT Instrument

Respondent A1 indicated that “it is very valuable for comparing of e-assessment tools”. Respondent A2 added that “this instrument is valuable to apply within an HE environment especially for the overall university department which is responsible for choosing the online Computer Based Testing (CBT) software – this instrument can help the institution make a better choice of tool”. Supporting this, respondent A3 stated that she “would love the criteria of this framework to be given to the owners of the system I am using. I believe the system I am using is used in the banking training system and not suitable for universities”. Respondent A4 found that the SEAT Framework was “easy to use but requires some degree of thought”. He added that “the length of the instrument is essential and definitely not off-putting”. These are rewarding remarks, since the use of SEAT for evaluating e-assessment tools to be adopted by academics, is one of the researchers’ main intentions.

Applicability of the SEAT Instrument

All comments regarding the applicability of this instrument were positive, “… most applicable” said A1. A3 indicated that “this framework could be used to show to non-users of a system, the wonderful features of a system”. A4 found the framework “easy to implement and easy to administer... non-intrusive”.

Value of this Instrument

Respondent A3 stated that the SEAT Framework is "most useful when considering the purchase of a system". A4 suggested that it provides "a comprehensive overview of the most important features of an e-assessment tool". A1's comments supported this, indicating that "all the relevant questions about an e-assessment tool are already in the Instrument. You can just answer the questions to evaluate the tool". Especially "within a teaching environment, the value of this instrument lies in empowering the user to make a better choice between various CBT software packages," stated A2.

Benefits of this Instrument

The practical benefits of the SEAT Framework outlined by the participants included "saves you time when comparing available e-assessment tools" (A1). A2 indicated that it "can be used as a bench-marking instrument to benchmark various online assessment tools – no such benchmarking tools exist within a SA environment"; A3 felt that it is "most useful when considering the purchase of a system"; and A4 believed that SEAT is "one of the few comprehensive tools available".

Possible Shortcomings/Gaps within this Instrument

In the open-ended responses, none of the participants identified shortcomings in the SEAT Framework – "Can't think of something," said A1 and A4 confirmed this saying, "none that I could identify". Furthermore, no participants identified any criteria which needed greater clarity when they were investigating their e-assessment tool.

These affirmations indicate the successful evolution of the framework.

Subsequent to Study 4d, no criteria were deleted from the Application Framework. However, two criteria were identified for inclusion in the electronic version of SEAT (e-SEAT). These were criteria 10 and 11 in the *Technical Support* category (Appendix J1). This resulted in e-SEAT consisting of eleven categories, and 182 criteria (see Table 7.2).

6.1.5 Conclusion of Study 4

In Study 4 of this action research series, a comprehensive framework, termed SEAT, was developed to evaluate e-assessment systems. This was achieved through four substudies each of which improved the previous versions of the framework, which evolved from the Pilot Framework in Study 4a, to the Evaluation Framework in Study 4b, then to the Proof of Concept Framework in Study 4c, and finally to the Application Framework in Study 4d, as shown in Figure 6.1 and Figure 6.2.

Study 4a (Section 6.1.1) was used to obtain essential feedback on the design, content and validity of the framework in this research. Study 4b (Section 6.1.2) determined which of the criteria identified in Phase 1, were essential for any e-assessment tool and which should be omitted or reworded. Study 4c (Section 6.1.3) provided further insight into the essential criteria to be included in the SEAT Framework. Finally, Study 4d (Section 6.1.4) took the SEAT Framework to participants who were experienced in using e-assessment. They were required to try it out in an authentic situation, each investigating its application to an existing/adopted e-assessment system.

At the end of Study 4, the researcher felt it necessary to create a scoring system for the criteria identified in the SEAT Framework, so that a tangible summarised and quantified report could be provided to the user, after evaluating an e-assessment tool or system using SEAT Framework. This meant that an electronic version of SEAT should be developed that, among various other rich features, could perform calculations.

Hence, together with responses to the suggestions made in the qualitative comments provided by participants in the various substudies of Study 4, as shown in Tables 6.3, 6.17 and 6.22, a scoring and totalling system, was introduced into the electronic version, the e-Seat (electronically Selecting and Evaluating an e-Assessment Tool) Framework, developed for Study 5.

6.2 Study 5: e-SEAT Framework evaluation

Prior to Study 5, the SEAT Framework that was developed over the four iterations of Study 4, was converted to an electronic version, to be called the e-SEAT Framework. Its purpose was to facilitate automated evaluation and rating of existing e-assessment systems or adoption of new systems. Moreover, the comprehensive report provided by e-SEAT provides the user with a clear evaluation of the features present, as well as those lacking, in the e-assessment tool evaluated by them, through the e-SEAT Framework. To facilitate this, a programmer was employed to generate this e-Framework, implementing the researcher's design. e-SEAT, with its eleven categories and 180 criteria (see Table 7.2), was then hosted on the UKZN server. Study 5 is outlined in Table 6.24 (similar to Table 4.9 in Section 4.5.5).

Through the process of e-SEAT evaluation, Study 5 contributed to answering Research Question 6.

Research Question 6	*How appropriate and effective is the proposed framework?*

Table 6.24: Summary of Study 5 as outlined in Figure 6.1

Study 5 (Oct – Nov 2012) - e-SEAT Framework Evaluation	
Respondents	4 expert users (UKZN, DUT, UP and WITS)
Data Collection	Electronic Instrument e-SEAT (See Appendix J1), Questionnaire (See Appendix J3) and Follow-up Interviews, where required
Data Analysis	Quantitative – basic statistical analysis Qualitative – content and discourse analysis
Purpose	To evaluate the electronic version of SEAT (called e-SEAT – electronically Selecting and Evaluating an e-Assessment Tool)

Following the initial creation of e-SEAT (Appendix J1), this framework, in turn, needed evaluation and refinement in line with the action research approach. It must be noted that the version of e-SEAT provided in Appendix J1 and referenced in Table 6.24, is the final e-SEAT that was produced at the end of Study 6. Since e-SEAT was designed as an online interactive framework, all changes that occurred in the process of evaluating the e-SEAT Framework, resulted in e-SEAT overwriting itself.

Hence the intermediate versions that were input to Study 5 and that resulted at the end of Study 5, were not stored and are not available to display.

A summary of the changes which were implemented on e-SEAT following this Evaluation Study (Study 5) is presented in Table 6.26.

In Study 5, four expert users of e-assessments, who had not participated in any of the iterations in Study 4, were invited to participate in the Study, that is, it was a purposive sample. Participants were initially required to use the electronic framework, e-SEAT, as a framework to evaluate an e-assessment system they were currently using, or had used in the past, or were considering using in the future. Thereafter they evaluated e-SEAT.

Following their interaction with e-SEAT, they were required to answer a questionnaire based on their experience of assessing an e-assessment tool/system using e-SEAT, followed by an interview. The follow-up interview had no set questions, but was conducted to confirm aspects in the qualitative comments that required further clarification. One participant (ESP5) did not provide any qualitative comments and was unavailable for a follow-up interview, hence was excluded from this dataset. To facilitate their interaction, participants were provided with the following:

- the link where the e-SEAT Framework is being hosted – http://major.ist.ukzn.ac.za/upasana/SEATVer1/Home.aspx ,
- a brief instruction file on how to use it (Appendix K), and
- a short questionnaire to answer, on their experience with the e-SEAT Framework (Appendix J3).

As depicted in Figure 6.6, participants reported positive interaction with e-SEAT, indicating that it is a useful framework, being intuitive to use and very few changes were required. In terms of usability, minimal features were lacking. There were a few suggestions for additional processing features and some shortcomings were mentioned, which are classified in Section 6.2.1.

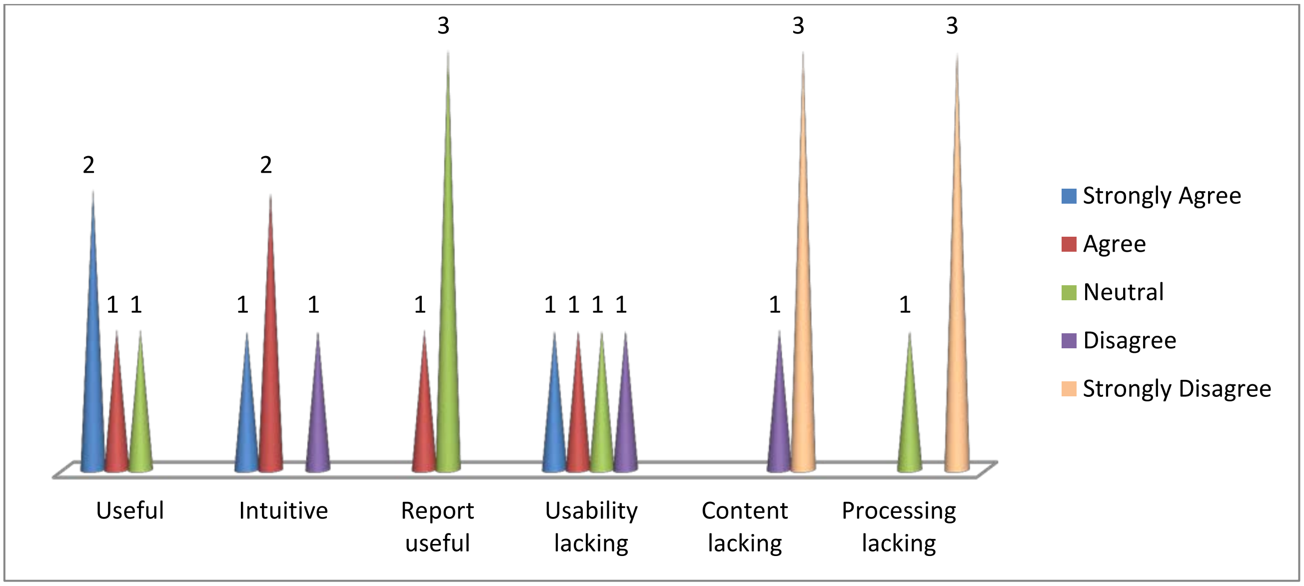


Figure 6.6: Reponses on interaction with e-SEAT: e-SEAT evaluation study

6.2.1 Discussion of findings

This section integrates the qualitative responses to the questionnaire and the interviews. Nine themes emerged and are listed in Table 6.25. The questionnaire also included questions on use of the e-SEAT instruction file and additional features the participants might require.

Table 6.25: Aspects discussed in Study 5

	e-Seat Themes
1	Usefulness
2	Intuitiveness
3	Report usefulness
4	Usability features
5	Content features
6	Processing features
7	Benefits
8	Positive features
9	Negative features

6.2.1.1 e-SEAT usefulness

Respondent ESP2 commented that “e-SEAT is very useful to evaluate my e-assessment tool as it covered most aspects that one requires in such a tool”. This indicated that e-SEAT was a comprehensive framework.

6.2.1.2 e-SEAT intuitiveness

Three participants indicated that e-SEAT was intuitive to use. ESP3 found that “presenting it in a survey type of thing it was easy to use”. However, ESP2 proposed that having “a clearer indication of your progress could potentially enhance the tool”. This was also suggested by ESP4 who stated that “it might be beneficial to show the user clearly how much of the instrument he/she has completed and how much is still outstanding”. However, ESP2 added “... but if you read the instructions you can work out your progress”.

6.2.1.3 e-SEAT report usefulness

ESP2 found it useful, stating that the report provided her “with useful information about the tool I am evaluating”.

6.2.1.4 e-SEAT usability features

Usability features that needed improvement included:

- The ‘Print Results’ button on the reports page also opens an email which is confusing” (ESP2).
- “It would be more useful if it informed you in a simpler way as to what section you were doing, and how far you had to go” (ESP3).
- He also commented on the length of the framework suggesting, that “there could be fewer questions – but then I suppose it depends on how detailed a review a person wants”.

6.2.1.5 e-SEAT content features

ESP2 stated that e-SEAT offers “very comprehensive coverage of the aspects to consider when evaluating an e-assessment tool”. ESP3 felt that e-SEAT should have two kinds of features, namely, in-built features of the e-tool and features that are customisable or add-ons during installation.

6.2.1.6 e-SEAT processing features

Some suggestions for additional processing features were given by the participants. ESP4 felt that the absence of a “Not sure option … could maybe force an incorrect answer”, if the user did not know the full features of the tool. ESP2 found “offline and online not clear, as my tool is also downloaded to my computer. As the tools allow you to work on any assessment then ask if you want to publish/have it available to users. Is publishing on the server the online part?” He also added that “some of the features have to be activated. The instrument assumes them to be standard features, for example, flagging of answered questions; allowing multiple editors”. According to ESP4, “the inability to cancel an option” was problematic. He further added that “once clicked you must provide a response to that question, even if you would rather leave the question unanswered, after realising that it does not adequately capture your answer”.

6.2.1.7 Benefits of e-SEAT

Participants indicated a number of people who would benefit from using e-SEAT. ESP1 proposed that it would be helpful to non-adopters of e-assessment who are, nevertheless, potential users, “academics who intend on using e-assessment in future”. ESP2 said that "people who have to choose between different e-assessment tool options or who might want to evaluate their existing e-assessment tool would benefit significantly by using this tool. SEAT would highlight positive aspects of the tool being evaluated, as well as point out missing features”. ESP3 believed that decision-makers would benefit greatly if they had previously “worked with them (e-assessment) quite a lot, or if they have extensive knowledge of the tool. This is due to the depth and range of questions in the framework”. ESP4 pointed out that an assessor who is planning to use an e-assessment tool in the future, and who does not fully know the features of e-assessment systems, might not be able to appropriately assess the e-assessment tool being investigated, without a tool such as e-SEAT.

6.2.1.8 Positive features, aspects and functions of e-SEAT

ESP1 appreciated that “technical support features were addressed” while ESP2 had experienced that “the comprehensive coverage of SEAT is outstanding!” Both ESP2 and ESP3 indicated they were exposed to more features of e-assessment than they had been aware of, through e-SEAT, “… it gets you to realise there are so many things that make up e-assessment”, said ESP3. ESP2 further pointed out that “e- SEAT also prompted me to investigate more about my tool in areas where I was unsure whether the tool had such options”. ESP4 stated that “the fact that it is an evaluation tool is a good thing … I think it can be very helpful”.

6.2.1.9 Negative features of e-SEAT

The length of the framework was a concern to ESP1, “So many questions. But necessary of course”. The issue of not knowing the progress made in the framework was also raised by ESP3, who stated “… not knowing where I was in the ‘quiz’”. She also added that “… having the ‘Calculate’ and ‘Clear Page’ buttons so close together – I could easily hit the ‘Clear’ button and have to re-enter the page again”

6.2.1.10 e-SEAT Instruction File

Three of the four participants were able to use e-SEAT without referring to the instruction file. Although ESP2 did use the instruction file, she indicated minimal use of it, stating that “I kept the instruction file open, so after registering I referred to it, and then at the end, I remembered I must do something, so I referred to the file to confirm that I first had to print a local copy”. Furthermore, all the participants found the instruction file “user friendly” and “short”.

6.2.1.11 e-SEAT Additional Features recommended

Participant ESP1 suggested a useful feature to include “… to automatically email the results to the user”. This would help inadvertent loss of the results if a user “clicks the ‘Close’ button”.

Table 6.26 summarises the changes suggested by the participants in Study 5, and the actions taken by the researcher to create the next version of e-SEAT, namely, the Validation Framework.

Table 6.26: Summary of changes made to e-SEAT based on the results of Study 5

e-Seat themes	Suggestions by participants in e-SEAT Evaluation Study	Participant code	Researcher's response
Intuitiveness	Indicate progress within the framework	ESP2, ESP3, ESP4	Inclusion of a progress bar was attempted by the programmer who had done the technical development of the electronic framework, e-SEAT
Usability features	Print Results button confusing	ESP2	Tasks separated in report – one for printing, the other for emailing
	Decrease length of framework	ESP3	All criteria deemed necessary, as the framework is meant to be comprehensive
Content features	Indicate custom features	ESP3	Further subcategorisation carried out by the researcher had resulted in the framework being over-complex. Hence a trade-off was made in favour of usability
Processing features	Include a "not sure" rating	ESP4	Not required, as the N/A option is available. The instructions were modified to indicate this
	'Offline' and 'online' terms confusing	ESP4	These criteria were reworded where appropriate
	Inability to cancel an option is problematic	ESP4	A feature that permits the user to cancel an option is not recommended by the developer, because standard interface design principles have been adhered to. Hence, the instruction wording has been modified and a clear message is provided if any criterion has not been rated
	Email report to users	ESP1	An email feature was incorporated by the developer
Negative features	Calculate and clear page button positioning	ESP3	These were moved for better usability

6.2.2 Conclusion of Study 5

Study 5 provided a partial answer to Research Question 6 'How appropriate and effective is the proposed framework?' because Study 6 also contributes to answering this research question. To determine the effectiveness of the framework four expert users of e-assessments, who had not participated in any of the iterations of Study 4, were invited to critically evaluate e-SEAT. This brought an unbiased perspective to the investigation of e-SEAT. Participants used the electronic framework (e-SEAT Evaluation Framework as outlined in Figure 6.2 in Section 6.1) to evaluate an e-assessment system they were using, or had used in the past, or were considering for the future. Thereafter, based on their experience with e-SEAT, they completed a brief questionnaire. The purpose of the questionnaire (Appendix J3) was to capture the essence of their interaction with the e-SEAT Evaluation Framework, as well as give them an opportunity to make suggestions for further improvements. These changes were implemented, generating the e-SEAT the next version of e-SEAT, which was validated in Study 6.

6.3 Study 6: e-SEAT Framework application and validation

Following Study 5, the e-SEAT Evaluation Framework was adapted, based on suggestions of the participants in Study 5 (Table 6.26). This resulted in the e-SEAT Validation Framework, as shown in Figure 6.2. Thereafter three users who are specialists in the use of e-assessment were contacted to participate in Study 6. As presented in Table 6.27, which is based on Table 4.10 in Section 4.5.6, the objective of this study was primarily to validate the e-SEAT Validation Framework. A secondary aim of Study 6 was the application, by each participant, of the e-SEAT Validation Framework, to an e-assessment tool they used. Through the process of validation and application of the e-SEAT Framework, Study 6 contributed to answering Research Question 6.

Research Question 6	*How appropriate and effective is the proposed framework?*

Table 6.27: Summary of Study 6 as outlined in Figure 6.1

Study 6 (May – July 2013) – e-SEAT Framework Application and Validation	
Respondents	3 expert users (CPUT, MEDUNSA, UJ)
Data Collection	Electronic Instrument (See Appendix J1), Questionnaire (See Appendix J3) and Follow-up Interviews, where required
Data Analysis	Qualitative – content and discourse analysis
Purpose	To apply and validate the FINAL electronic framework developed (e-SEAT)

In purposive sampling, participants were carefully selected, one being a developer of an e-assessment tool; the second being an experienced user of both MCQs and e-assessment; and the third being both an expert user as well as a prolific researcher in the area of e-assessment. They were required to critically review the e-SEAT Framework and validate it. They were also asked to use e-SEAT to evaluate an e-assessment system they used previously. Finally, they were required to answer a brief questionnaire based on their experience of interaction with e-SEAT. Participants were provided with the same material as given to the participants in Study 5, i.e.

- the link where the e-SEAT Framework is being hosted – http://major.ist.ukzn.ac.za/upasana/SEATVer1/Home.aspx ,
- a brief instruction file on how to use it (Appendix K), and
- a short questionnaire on their experience with the e-SEAT Framework (Appendix J3).

As explained in Study 5, the version of e-SEAT provided in Appendix J1 refers to the final e-SEAT produced at the end of Study 6. Since e-SEAT was designed as an online interactive framework, all changes resulted in e-SEAT overwriting itself. Hence the intermediate version that emerged after the end of Study 5, was not stored and is not available to display.

A summary of the changes which were implemented on e-SEAT following this Application Study (Study 6) is presented in Table 6.29.

As illustrated in Figure 6.7, participants in Study 6 reported more positive interaction with e-SEAT than the participants in Study 5. Their responses confirmed that e-SEAT is both useful and intuitive to use, thus contributing to validating it. A suggestion was made on improving the usability of e-SEAT, which is discussed in Section 6.3.1. Very few additional processing or content features were recommended. In contrast to Study 5, all participants found the report that e-SEAT generated, to be useful for taking decisions on adoption.

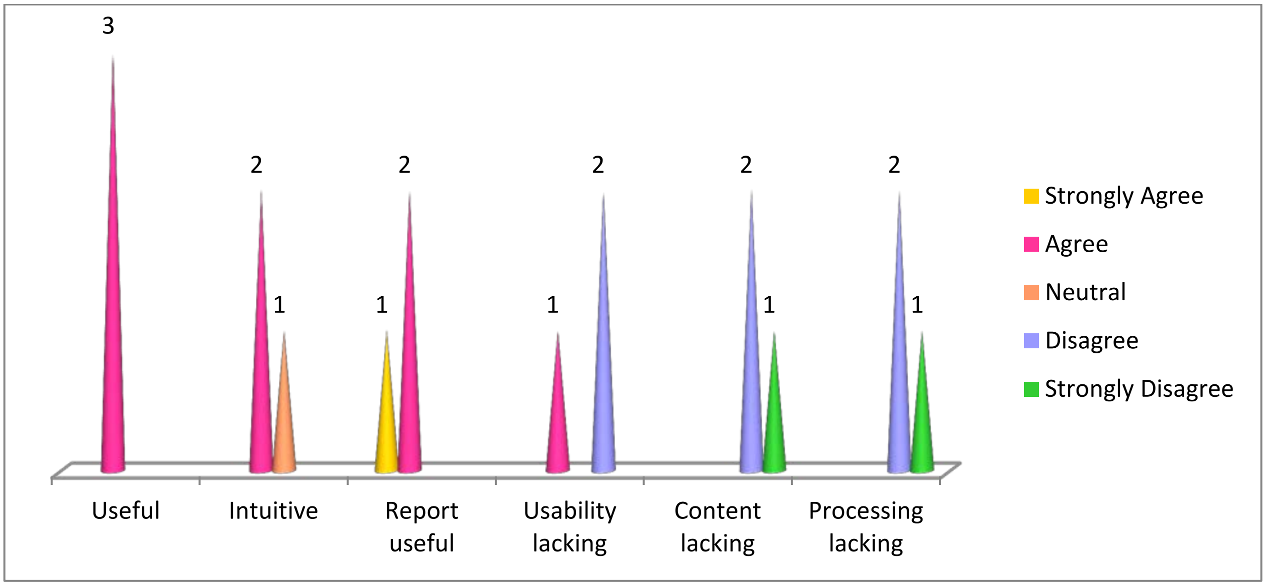


Figure 6.7: Reponses on interaction with e-SEAT: e-SEAT validation study

Some of the qualitative aspects of Study 6 are discussed in more detail in Section 6.3.1.

6.3.1 Discussion of findings

The same themes that were used in Study 5 and that are listed in Table 6.28, were also discussed in Study 6, as listed in Table 6.28. However, qualitative comments were only provided for some of the themes, namely those highlighted in blue. These are discussed in Sections 6.3.1.1 to 6.3.1.7, together with mention of the instruction file and additional comments.

Table 6.28: Aspects discussed in Study 6

	e-Seat Themes
1	Usefulness
2	Intuitiveness
3	Report usefulness
4	Usability features
5	Content features
6	Processing features
7	Benefits
8	Positive features
9	Negative features

6.3.1.1 e-SEAT intuitiveness

Two aspects relating to intuitiveness were commented on by VSP1: "… tricky to know where you are" and "… the green text indicating where one is in the framework could be more visible!"

6.3.1.2 e-SEAT usability features

Once again, the issue of orientation occurred. VSP1 found the aspect of "knowing where you are" problematic. This concern had also been raised by participants ESP2, ESP3 and ESP4 in Study 5.

6.3.1.3 Benefits of e-SEAT

Among the type of stakeholders who would benefit from e-SEAT, participants suggested "administrators, budget people" (VSP1) and academics using the electronic assessments (VSP2 and VSP3). Similar to what ESP2 indicated in Study 5, was VSP3's suggestion "….. that you can only assess a tool once you know it well". Hence, he suggested that "a database of assessments that were done by users knowing the tool well" be compiled from the results of e-SEAT and "people that want to buy an assessment tool can go and look at assessments (done by independent people) and then decide what tools they must consider". He added that, at times, "people with very little knowledge of assessment tools have to make decisions that involve lots of money. So before they buy a tool they must have access to assessments of the different tools".

6.3.1.4 Positive features, aspects and functions of e-SEAT

Participants valued the comprehensive list of criteria. It not only showed "possible features" but it also "lists things one has not even thought of!" said VSP1. VSP3 pointed out that "a tool may have a low score in a certain feature but that feature may not be important to you so you can ignore the low score", hence the further categorisation into essential and optional criteria may be necessary prior to the future implementation of e-SEAT.

6.3.1.5 Negative features of e-SEAT

VSP1 disliked "… the full justification of the questions". She further added that "it took me a while to realise where I was", reiterating the need for a progress indication in the electronic framework. VSP2 suggested that "the instructions for clicking on different aspects of SEAT could have been more obvious, and not in a separate document", indicating the need for online context-sensitive help. VSP3 reiterated the need for an option to "allow the users to choose the features that they want the final score to be calculated on, and omit certain features when the final score is calculated", which is a highly pertinent contribution.

6.3.1.6 e-SEAT Instruction File

All participants felt that e-SEAT was sufficiently intuitive to be used without the instruction file. VSP1 and VSP2 suggested that the Iength of the instruction file could be reduced. VSP3 requested that the "weight of the features be included in the file".

6.3.1.7 e-SEAT Additional Comments

VSP1 proposed that e-SEAT should be used by "decision-makers as part of the decision-making process in making a new purchase". It is particularly useful because it "shows all possibilities of an assessment system", said VSP2. Thus, "by making evaluations of assessment tools available to people that have to decide between different tools, perhaps via a website" (VSP3), it will assist "people with very little knowledge of assessment tools" who "often have to make decisions that involve lots of money", added VSP3. These comments demonstrate the utility of e-SEAT in supporting lateral thinking by users who are evaluating an e-assessment system or considering the adoption of one. e-SEAT prompts them to consider aspects they might not have thought of independently.

VSP3 believed that "there are certain areas where a tool must score close to 100% otherwise the tool is useless, for example, security and reliability, interface and the results. Test Results are the feedback that you get from your course, which will help you to improve your course and teaching".

VSP1 reiterated the need for a progress bar, since the framework is very lengthy and users could get frustrated not knowing where they are in the completion process. Alternatively, she suggested that the headers on each page indicating which Category was being evaluated should be made more prominent – "BOLD, RED, LARGE font" – since she was sometimes confused as to what she was evaluating. She suggested omitting some of the preamble at the top that is repeated, since the questions or criteria only start one third of the way down the page. Finally, she added that the "pdf Help and Instruction file should be made more visible".

Table 6.29 summarises the changes suggested by the participants in Study 6, and the actions taken by the researcher.

Table 6.29: Summary of changes made to e-SEAT Validation Framework based on the results of Study 6

e-Seat themes	Suggestions by participants in e-SEAT Validation Study	Participant Code	Researcher's response
Intuitiveness	Indicate progress within the framework	VSP1	Inclusion of progress bar attempted by the programmer, who had done the technical development and coding of e-SEAT. However, he was unsuccessful in implementing this in the current version of e-SEAT. The researcher will include it in the next version of e-SEAT, subsequent to the PhD study
	Instructions unclear	VSP1	Wording and font colour of instructions revised, made more understandable and visible
Processing features	Further subcategorisation into essential and optional criteria	VSP3	Not feasible for this current implementation of e-SEAT. The researcher will review this and include it in the next version of e-SEAT, subsequent to the PhD study
Negative features	Full justification of the questions	VSP1	Question alignment adjusted to left-alignment for better readability
	No online context-sensitive help	VSP2	The researcher will review the need for this and possibly include this in the final version of e-SEAT, subsequent to the PhD study
	Length of the Instruction File could be reduced	VSP1, VSP2	The Instruction File was substantially reworked
	Weight of the features should be included in the Instruction File	VSP3	Criteria weighting included in the Instruction File
	Headers not distinct enough	VSP1	Headers made more prominent
	Instruction file not visible	VSP1	Help Tab created

6.3.2 Categories and criteria in e-SEAT Final Framework

At the conclusion of the six action research studies that comprise Phase 2, the creation and refinement of SEAT and e-SEAT, this section presents the complete list of categories (11) and criteria (182) that are included in the e-SEAT Final Framework (Appendix J1). Tables 6.30 to 6.36 provide the 'Functional' criteria, while Tables 6.37 to 6.42 list the 'Non-Functional' criteria included in the Final e-SEAT Framework. For improved readability, Tables 6.30 and 6.31; 6.33 and 6.34 depict, in parts, the full list of criteria found respectively in the *Question Editing* and *Test and Response Analysis* categories of the Final e-SEAT Framework.

The 31 criteria that are included in the *Question Editing* category of the Final e-SEAT Framework, are provided in Tables 6.30 and 6.31.

Table 6.30: Question Editing category in e-SEAT

Question Editing - The Software:	
1. allows the academic to create the test electronically	11. facilitates offline question creation within the tools
2. updates the test bank immediately, and not at the end of the session, when questions are edited/authored	12. grants academics previews of assessments created offline
3. permits the academic to author original questions to add to the question bank	13. grants academics previews of assessments created online
4. allows the academic to view existing questions in the question bank	14. incorporates an automatic grammar check facility
5. allows the academic to adapt existing questions in the question bank	15. incorporates a spell checker
6. supports importing of questions in nonproprietary, interoperable format to the question bank	16. flags questions which learners have not answered in an assessment, so that these can be deleted or amended by the academic
7. supports exporting of questions in nonproprietary, interoperable format from the question bank	17. allows the academic to add comments to a question created before adding to/rejecting from the question bank where multiple editors are working on one test bank
8. permits a range of parameters/options to be specified in questions (e.g. four or five options per question)	18. allows the academic to approve or reject all created questions before adding to/rejecting from the question bank
9. supports feedback creation for each question	19. directs comments regarding questions submitted to the question bank directly to the author of the question
10. allows the incorporation of question metadata (e.g. categories, keywords, learning objectives, and levels of difficulty)	20. allows academics to create a marking scheme for an assessment
	21. allows academics to combine questions from different test banks into a single test

Table 6.31: Question Editing category in e-SEAT continued …

22. allows academics to pilot tests prior to the assessment going live
23. supports printing of tests for moderating purposes
24. records average time taken by learners for each question
25. facilitates allocation of marks to questions to support manual marking
26. provides support for incorporating graphics in questions
27. provides tools to do automatic analysis of learner responses
28. supports printing of tests to support taking of the test offline
29. facilitates allocation of marks to questions to support overriding the mark automatically assigned
30. flags questions as easy, average or difficult (metadata) to support better randomisation
31. displays the IP address of the individual learner taking the test

The eleven (11) criteria that are included in the Assessment Strategy category of the Final e-SEAT Framework, are provided in Table 6.32.

Table 6.32: Assessment Strategy category in e-SEAT

Assessment Strategy - The Software:
1. supports random generation of questions from the test bank in multiple versions of the same assessment
2. incorporates branching of questions, depending on the learners' response (e.g. if the learner selects option (a) questions 5 to 10 are displayed, else questions 11 to 15 are displayed)
3. displays feedback as/if required
4. displays results as/if required
5. specifies how many attempts a learner is permitted to make on a question
6. permits learners to sit a test as many times as they like, in the case of selfassessments
7. permits a learner to take the test at different times for different sections, in the case of selfassessments (e.g. complete section A today, section B tomorrow and eventually complete paper when he/she has the time)
8. permits learner to take a self-assessment offline
9. supports test templates that facilitate many types of testing including formative, peergenerated, practice, diagnostic, pre/post and masterylevel testingly
10. automatically prompts learners to redo an assessment (with different questions covering the same topics) if they get below a specified percentage
11. automatically prompts learners to redo an assessment (with questions they previously answered correctly removed from the new assessment) if they get below a specified percentage

The 36 criteria that are included in the *Test and Response* category of the Final e-SEAT Framework, are provided in Tables 6.33 and 6.34.

Table 6.33: Test and Response category in e-SEAT

Test and Response - The Tool:	
1. allows groups to be set up	10. presents results immediately to learners, when appropriate
2. allows learners to be added to a group	11. provides learners with the option/facility to print out assessment responses
3. permits questions to be viewed by metadata fields (e.g. categories, keywords, learning objectives, and levels of difficulty)	12. distributes academics' comments to learners via the system
4. allows learners access to previous assessment results	13. distributes academics' comments to learners via email
5. allows learners access to previous assessment responses	14. emails academics automatically if the marking deadline is not met
6. allows learners access to markers' comments on prior assessments (in cases where a human assessor reviewed the completed test)	15. presents mean (average) score statistical analysis per assessment
7. allows results to be accessed after a specific date, as required	16. presents discrimination index statistical analysis per assessment
8. allows learners to compare the results they obtained with other learners' results	17. presents facility index statistical analysis per assessment
9. allows learners to compare marks with group averages	18. presents highest score statistical analysis per assessment
	19. presents lowest score statistical analysis per assessment
	20. presents frequency distribution statistical analysis per assessment

Table 6.34: Test and Response category in e-SEAT continued …

21. incorporates an automated 'cheating spotter' facility	32. the statistical analysis per assessment presents the percentage correct
22. supports the ordering of the results tables in various ways (e.g. by marks, student numbers, names, etc.)	33. the statistical analysis per assessment presents the difficulty index statistical
23. displays marks as percentages	34. the statistical analysis per assessment presents the percentage answered correct
24. presents, to the academic, all attempts at a question	35. the statistical analysis per assessment presents the percentage of top learners who got the question correct
25. permits the academic to view individual responses to questions	36. the statistical analysis per assessment supports correlation of assessment data across different class groups
26. allows the learner to view the whole test, as he/she had completed it	
27. displays a comparison of mark data of different groups	
28. displays a comparison of the performance in different subtopics/sections	
29. permits mark data to be viewed without having access to names of learners	
30. flags questions which were poorly answered	
31. flags questions which were well answered	

The four (4) criteria that are included in the *Test Bank* category of the Final e-SEAT Framework, are provided in Table 6.35.

Table 6.35: Test Bank category in e-SEAT

The Test Bank:
1. draws random questions from a question bank, as required
2. only contains questions which have been moderated for the required standard and cognitive levels
3. assigns global unique identifiers to all questions created or revised in the question bank
4. has the potential to include questions that test learners "Higher Order Thinking Skills" (HOTS)

The eighteen (18) criteria that are included in the *Question Types* category of the Final e-SEAT Framework, are provided in Table 6.36.

Table 6.36: Question Types category in e-SEAT

Question Types - The Tool Supports:	
1. Multiple choice: single response	9. Selection/Drop down lists
2. Multiple choice: multiple response	10. Ranking
3. True/false	11. Diagrams/Graphics
4. True/false with explanation	12. Video/Audio Clips
5. Fill-in-the-Blanks/ Completion	13. Drag and Drop
6. Simulation	14. Reordering/Rearrangement/Sequencing
7. Matching Items	15. Categorising
8. Extended Matching Items (EMIs)	16. Hotspots
	17. Hotspot (Drag and Drop)
	18. Text Input (short answer – which would be marked manually)

The fourteen (14) criteria that are included in the *Interface Design* category of the Final e-SEAT Framework, are provided in Table 6.37.

Table 6.37: Interface Design category in e-SEAT

Interface:	
1. is intuitive to use	9. provides an option to clearly display marks for each section
2. caters for users with special needs, by including features such as non-visual alternatives, font size variety, colour options	10. displays a clock to keep track of time allocated/ remaining for formative assessment
3. facilitates ways of varying the presentation of tests	11. allows academics to sms reminders to students of assessments due
4. allows learners to view all tests available to them	12. provides a toggle button to allow students the option to answer individual questions, or the whole assessment
5. permits learners to view logistical arrangements in advance, such as times and venues of assessments	13. presents help facilities for users
6. permits viewing of multiple windows as required for assessments	14. provides an option to allow/disallow printing
7. allows academics to email reminders to students of assessments due	
8. provides an option to clearly display marks for each question	

The 21 criteria that are included in the *Security* category of the Final e-SEAT Framework, are provided in Table 6.38.

Table 6.38: Security category in e-SEAT

Security Criteria - The Tool:
1. ensures that tests are accessible only to learners who have explicit authorisation, granted by access administrators
2. encrypts all data communicated via the network
3. ensures that mark data held on the server can be accessed by authorized persons only
4. logs the IP address where each learner sat
5. logs which questions were marked by which lecturer
6. logs when the academic marked the question
7. prevents answers to questions already completed from being altered (in cases where second opportunities are not permitted)
8. requires permission of the academic before any question can be modified or deleted from a test
9. prevents learners from amending a test once taken
10. prevents learners from deleting a test once taken
11. automatically allocates a global unique identifier to tests
12. provides the ability to view entire tests for verification without the ability to change them
13. restricts tests to particular IP addresses and domains
14. allows academics to enter details of learners who cheat to alert other colleagues of 'problematic' students
15. permits academics to modify results after communication with a learner regarding the reason for the change
16. permits test results to be changed or corrected when a memorandum error is discovered
17. logs modifications to original marks
18. records motivations for modifications to original marks
19. provides password access to tests
20. allows academics to restrict assessments to a specific IP address
21. prevents learners from opening any other windows not required for the assessment (similar to Respondus lockdown facility)

The eight (8) criteria that are included in the *Compatibility* category of the Final e-SEAT Framework, are provided in Table 6.39.

Table 6.39: Compatibility category in e-SEAT

Compatibility - The Tool:
1. is accessible from a standard, platform-independent web browser, without additional plugins
2. is downgradable for learners with previous versions of browsers
3. is customisable to provide a uniform interface with the rest of the institution's intranet or virtual learning environment
4. links seamlessly with other institutional systems, so that learners can use their existing username and passwords
5. permits results to be exported to spreadsheets or statistical analysis software
6. uses a common logon facility, that integrates with other institutional systems
7. links seamlessly with other institutional systems so academics can export marks directly
8. specifies which browser must be used for an assessment in the setup details

The 23 criteria that are included in the *Ease of Use* category of the Final e-SEAT Framework, are provided in Table 6.40.

Table 6.40: Ease of Use category in e-SEAT

Ease of Use - The System:
1. requires little time to capture data related to learner profiles and assessments
2. requires a short time period to set up an assessment online
3. requires little/no training on how to use the tool
4. provides simple and fast login procedures
5. includes an intelligent help system – dependent on the academic role and current activity
6. incorporates speech synthesis for 'special needs' learners
7. is intuitive to use – academics should not require any special programming language skills to adopt the tool
8. makes it easy to include multimedia elements in test items
9. allows academics access to details of times of an assessment
10. permits all learners in a group to be removed from the system simultaneously
11. allows access to details of learners sitting a test at a particular time
12. permits learners to return to the point at which they had exited an incomplete self-assessment
13. makes it easy,where necessary, to enter foreign characters and symbols
14. automatically distributes electronic certificates of test submission to learners
15. allows learners access to details of room numbers and venues of an assessment
16. allows learners access to details of times of an assessment
17. simplifies the task of adding learner access
18. simplifies the task of removing learner access
19. simplifies the task of editing learner access
20. allows learners to be enrolled on the system by an administrator
21. allows learners to be removed from the system
22. permits academics to enter learner details (name and student number) in the test directly
23. allows academics to limit a test by giving learners a unique number to access the test

The five (5) criteria that are included in the *Robustness* category of the Final e-SEAT Framework, are provided in Table 6.41.

Table 6.41: Robustness category in e-SEAT

Robustness - The Tool:
1. does not hang while a student takes a test
2. is stable, even when a large number of learners access the system or take a test simultaneously
3. does not crash frequently
4. is able to recover the test from the point at which the learner stopped, in the event of an unforeseen system error or crash
5. processes responses given by learners in an acceptable time period

The eleven (11) criteria that are included in the *Technical Support* category of the Final e-SEAT Framework, are provided in Table 6.42.

Table 6.42: Technical Support category in e-SEAT

Technical Support - The System:
1. incorporates a resilient network
2. if not web-based, includes software that is easy to install, requiring little effort and time
3. runs on multiple platforms
4. includes installation software that is easily available
5. allows new functionality to be incorporated without reinstalling the system
6. supports large numbers of concurrent learners are logged in simultaneously
7. supports multiformat data storage – Oracle/Access or ODBC (Open DataBase Connectivity) format
8. facilitates the use of existing database systems
9. grants academics access to details of all test purchases relevant to that academic, where tests are purchased from the supplier of the assessment software
10. automatically prompts learners to redo an assessment (with different questions covering the same topics) if they get below a specified percentage
11. automatically prompts learners to redo an assessment (with questions they previously answered correctly removed from the new assessment) if they get below a specified percentage

6.3.3 Conclusion of Study 6

Following Study 5, the e-SEAT Evaluation Framework was adapted, based on suggestions of the participants in Study 5 (Table 6.26). This resulted in the e-SEAT Validation Framework, as shown in Figure 6.2.

Study 6 completed the answer to Research Question 6 'How appropriate and effective is the proposed framework?' which was partially answered in Study 5. To achieve this, three specialist users in the field of e-assessment were contacted to participate in Study 6, the final study in the action research series to develop a real world product for evaluating e-assessment systems. As presented in Table 6.27, which is based on Table 4.10, the objective of this study was primarily to validate the e-SEAT Validation Framework. A secondary aim of Study 6 was the application of the e-SEAT Validation Framework to an e-assessment tool previously used by the participants.

Some of the suggestions made by these leading experts were incorporated into e-SEAT, which was changed as shown in Table 6.29. This led to an ultimate product called the e-SEAT Final Framework (Figure 6.2 and Section 6.3.2). Other suggestions were too major to be implemented at this stage of prototype development, but they have great potential, and will be considered by the researcher for inclusion in the next version of e-SEAT, subsequent to the PhD study.

The next few pages present a selection of screen prints of this final version of e-SEAT. These displays allow the reader to develop a good understanding of the electronic framework developed.

The framework is available electronically at
http://major.ist.ukzn.ac.za/upasana/SEATVer1/Home.aspx. Should the reader wish to personally utilise the e-SEAT Framework to evaluate an e-assessment system, it is recommended that this is completed anonymously, to preserve the integrity of the examination process. The e-mail that emerges on completion of an e-SEAT session will be received by an external third party and neither the student nor the supervisor will be aware of the transaction.

The conclusion to the chapter as a whole, follows after the screenprints.

6.4 Screen prints of e-SEAT

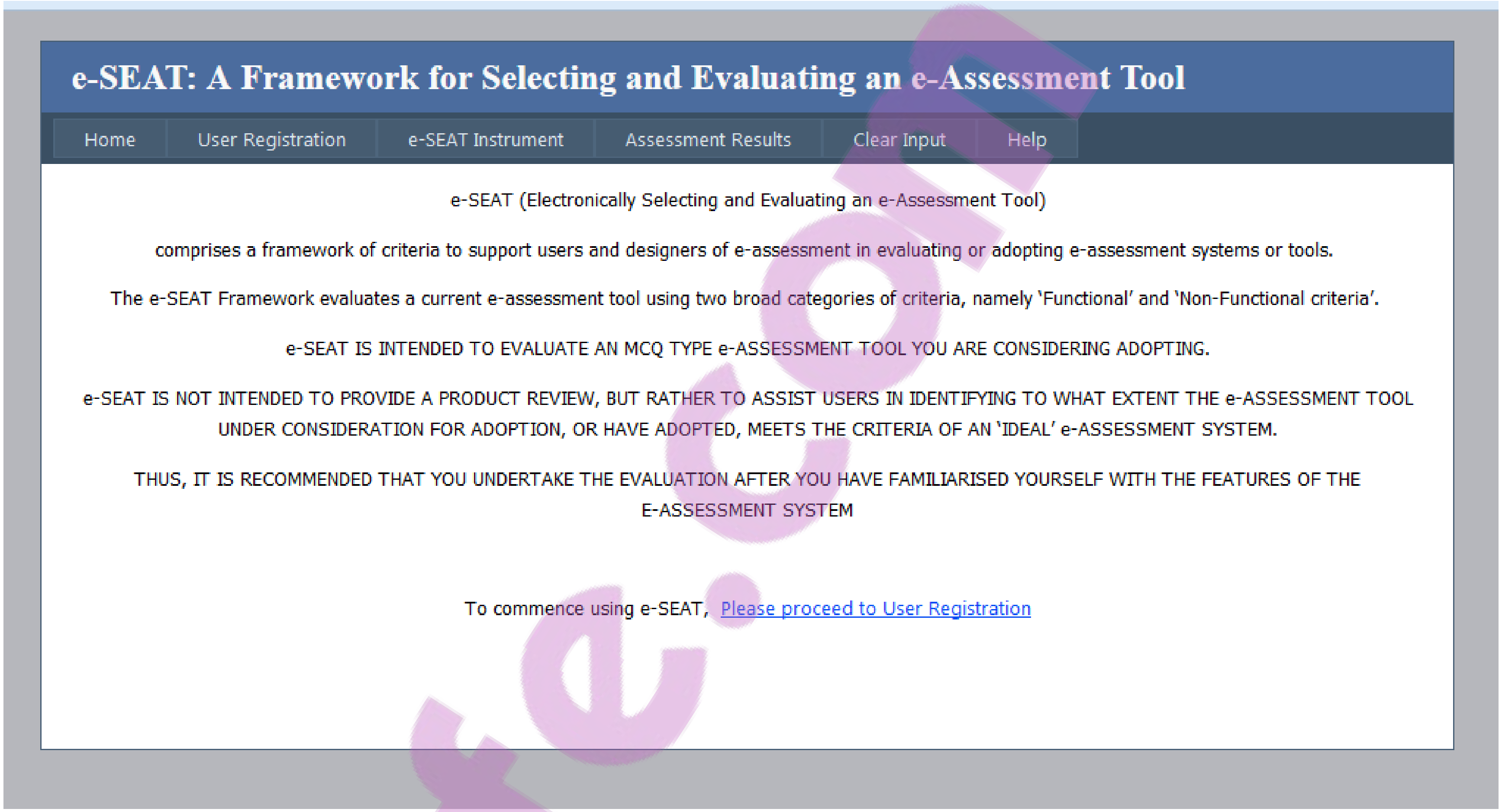


Figure 6.8: e-SEAT welcome screen

Figure 6.8 depicts the welcome screen that is displayed to the user when they access e-SEAT. It briefly outlines the purpose of e-SEAT.

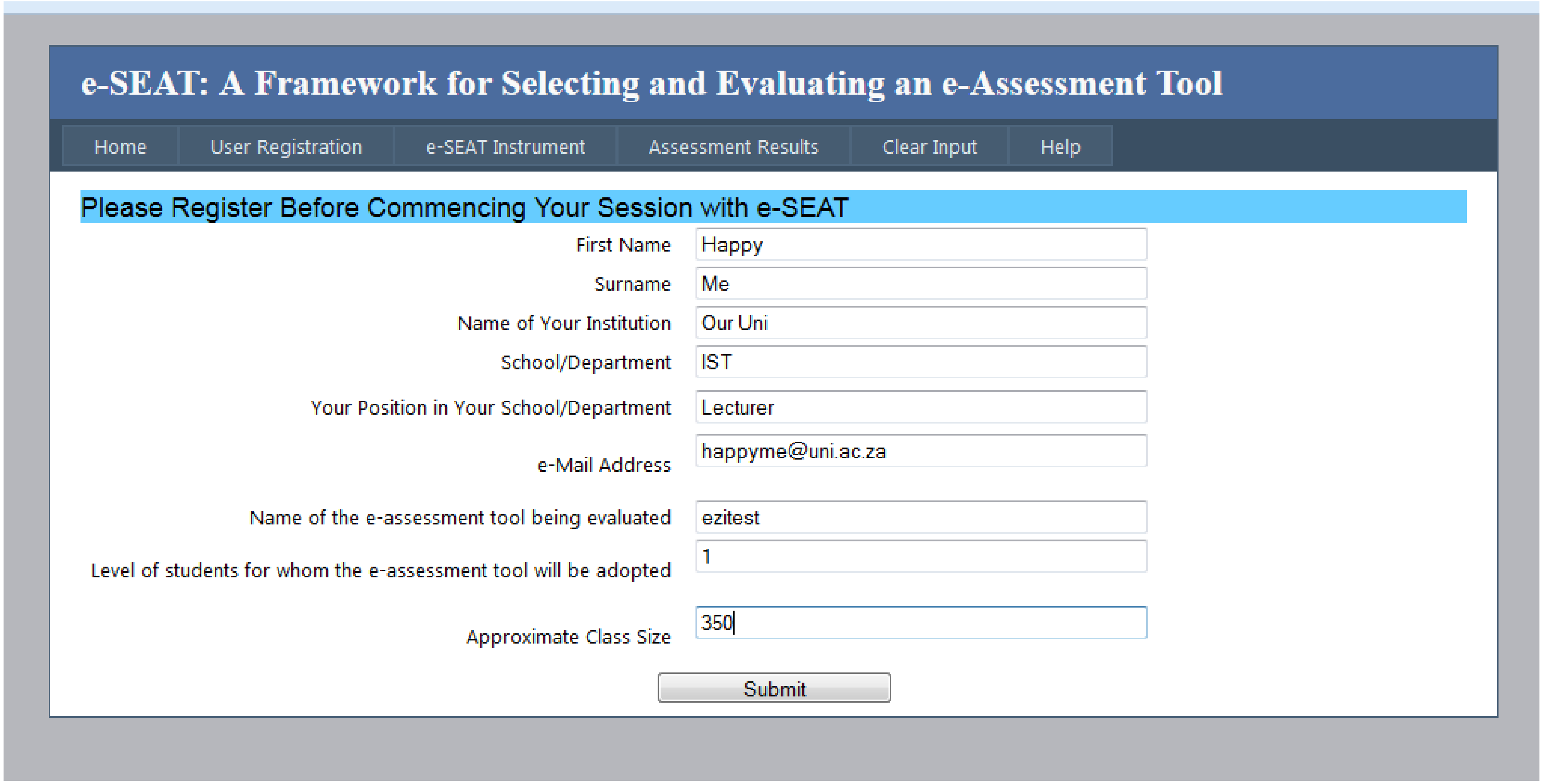


Figure 6.9: e-SEAT registration screen

In order to use e-SEAT to evaluate an e-assessment tool, the user is required to register first. The registration details required are presented in Figure 6.9.

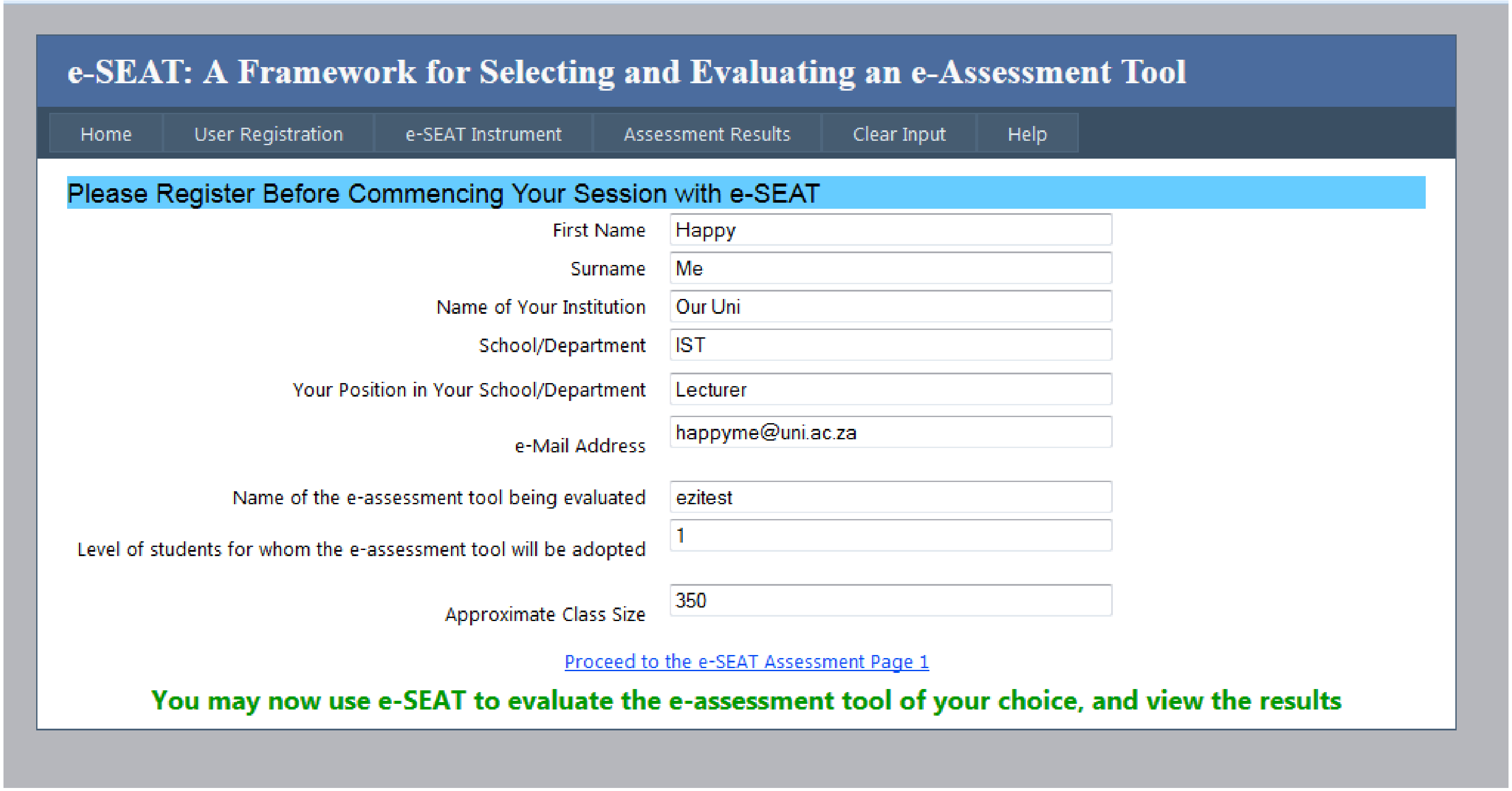


Figure 6.10: e-SEAT post registration screen

Post registration, the user is provided with a message as shown in green in Figure 6.10, informing him/her that the registration has been completed successfully, hence he may proceed to the first page of the e-SEAT Framework.

For improved readability, Figure 6.11, Figure 6.12 and Figure 6.13 depict, in three parts, the full screen that appears to the user for rating the criteria related to Question Editing. This supports the users in seeing how efficiently the e-assessment tool they are considering for adoption meets the requirements of an ideal e-assessment tool.

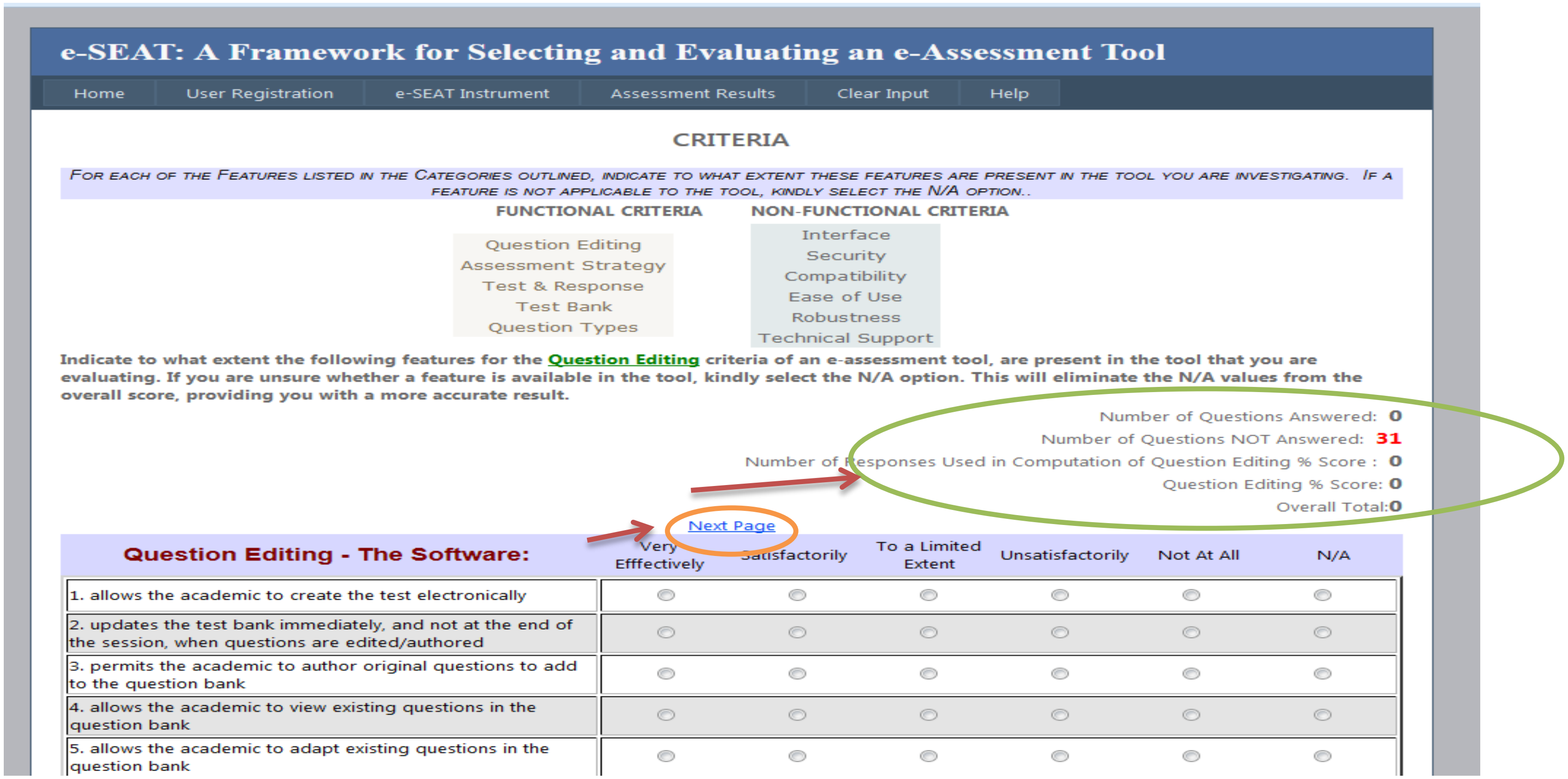


Figure 6.11: e-SEAT Question Editing Criteria screen

Information on the number of questions in the category, the number of questions unanswered, and the category as well as the overall score calculated, is presented on each screen, as highlighted in green in Figure 6.11.

6. supports importing of questions in nonproprietary, interoperable format to the question bank	○	○	○	○	○	○
7. supports exporting of questions in nonproprietary, interoperable format from the question bank	○	○	○	○	○	○
8. permits a range of parameters/options to be specified in questions (e.g. four or five options per question)	○	○	○	○	○	○
9. supports feedback creation for each question	○	○	○	○	○	○
10. allows the incorporation of question metadata (e.g. categories, keywords, learning objectives, and levels of difficulty)	○	○	○	○	○	○
11. facilitates offline question creation within the tools	○	○	○	○	○	○
12. grants academics previews of assessments created offline	○	○	○	○	○	○
13. grants academics previews of assessments created online	○	○	○	○	○	○
14. incorporates an automatic grammar check facility	○	○	○	○	○	○
15. incorporates a spell checker	○	○	○	○	○	○
16. flags questions which learners have not answered in an assessment, so that these can be deleted or amended by the academic	○	○	○	○	○	○
17. allows the academic to add comments to a question created before adding to/rejecting from the question bank where multiple editors are working on one test bank	○	○	○	○	○	○
18. allows the academic to approve or reject all created questions before adding to/rejecting from the question bank	○	○	○	○	○	○
19. directs comments regarding questions submitted to the question bank directly to the author of the question	○	○	○	○	○	○
20. allows academics to create a marking scheme for an assessment	○	○	○	○	○	○

Figure 6.12: e-SEAT Question Editing Criteria screen continued …

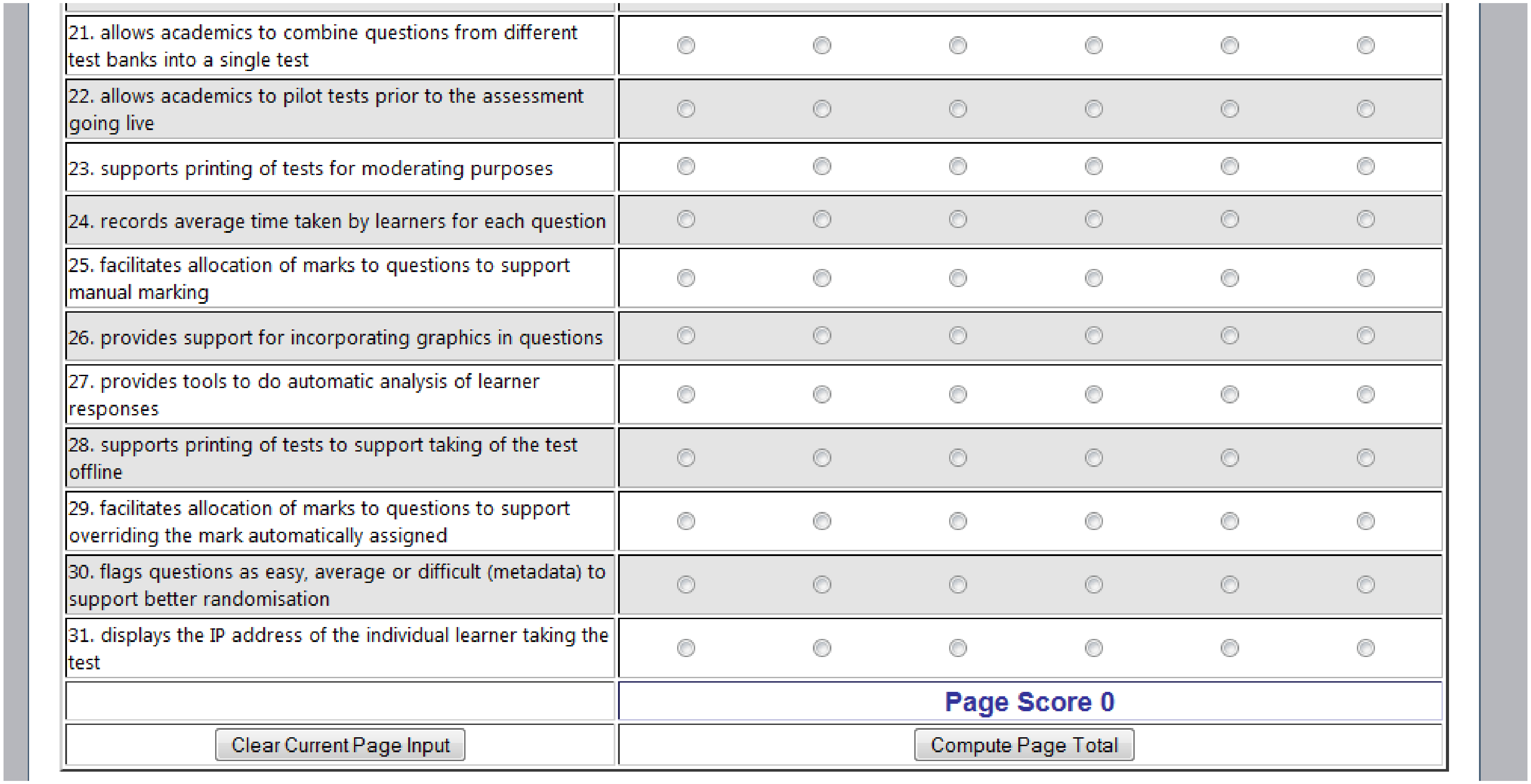


Figure 6.13: e-SEAT Question Editing Criteria screen continued …

Upon completion of the rating of all the criteria in the respective category, the Next Page link becomes available, as shown in orange in Figure 6.11. All eleven categories in e-SEAT, outlined in Table 6.17, present similar screens.

Upon completion of the evaluation of the e-assessment tool the user is considering adopting, e-SEAT presents a composite report to the user as illustrated in Figure 6.14. This report provides a summary of the score obtained in each category, as well as the overall score. The purpose of this report is to assist the user in his/her decision, whether or not to adopt the e-assessment tool under consideration.

e-SEAT Evaluation of your e-Assessment Tool

Total Number of Valid Responses (Used in Computation of e-SEAT %)	45
Total Number of Valid Functional Responses	20
Total Number of Valid Non-Functional Responses	25
Total Number of "Not Applicable"N/A Responses (Not used in Computation of e-SEAT%)	0
Total Number of UNANSWERED Questions (Not used in Computation of e-SEAT%)	137

FUNCTIONAL CRITERIA	
ASPECT	**FINAL % SCORE**
Question Editing	100
Assessment Strategy	100
Test Response Analysis	100
Test Bank	100
Question Types	100
NON-FUNCTIONAL CRITERIA	
Interface Design Component	50
Security	50
Compatibility	50
Ease of Use	50
Robustness	50
Technical Support	50
OVERALL RESULTS	
Functional Criteria	100
Non-Functional Criteria	50
Overall %	**72.22**

CONCLUDING REMARKS

Thank You **1 2** for using the **e-SEAT Framework** to evaluate the **7** e-Assessment Tool

The results of this Framework show that **100%** of **Functional Criteria** and **50%** of **Non-Functional Criteria** have been met by the

Figure 6.14: e-SEAT report screen

A copy of this report is emailed to the user as well, prior to him/her exiting the Framework. A print option is also available, should the user wish to print out the results.

The logic behind the calculation of the overall rating is as follows:

- The final %'s is calculated according to the number of questions answered in the Functional (F) and Non-Functional (NF) sections.
- In instances where the number of questions answered in the F section is more than the number of questions answered in the NF section, the % reported will be skewed in favour of the F section.
- In instances where the number of questions answered in the NF section is more than the number of questions answered in the F section, the % reported will be skewed in favour of the NF section.
- As an example, if there are 20 questions answered in the F section and 25 questions answered in the NF section, and let's suppose that the 20 F questions were all given a max score (i.e. "very effectively" which translates to a score of 4) and the 25 questions in the NF section were given a score of 2 (i.e. "to a limited extent") then the total of the F section would be 20* 4 which is 80 and the total for the NF section would be 25* 2 which is 50. The maximum possible score is 45* 4 = 180 (because there were a total of 45 questions answered – 20 F and 25 NF). The overall % is calculated by computing (80+50)/180. This gives an overall % of 72.22 (refer to attached image of the Results screen). Individually the F % is 100% and the NF % is 50%. If an average of this is computed, the answer will be 75% which is not an accurate reflection of the actual overall %.

6.5 Chapter conclusion

As presented in Figure 6.1, Phase 2 in this series of action research studies consisted of three studies, namely Studies 4 to 6. The initial version of the SEAT Framework, obtained from the criteria in Table 3.1 in Section 3.1.2 integrated with criteria from Table 5.35 in Section 5.4, served as the input for Study 4. Study 4 was iterative in nature, with four substudies, evolving from a Pilot Framework (Section 6.1.1) to the Evaluation Framework (Section 6.1.2), then to a Proof of Concept Framework (Section 6.1.3), and finally to an Application Framework (Section 6.1.4). Validation of the initial manual SEAT Framework occurred in the Application Study.

This rigorous sequence of evaluation and application facilitated the transition from the manual SEAT Framework to the interactive electronic framework, e-SEAT. Thereafter, Study 5 (Section 6.2) and Study 6 (Section 6.3) saw participants evaluating, applying and validating the electronic framework as e-SEAT progressed through its e-Seat Evaluation Framework and the Validation Framework. Thus the SEAT Framework and e-SEAT Framework were refined through a series of three main studies, one of them including four substudies, presented in this chapter. The Validation Framework, presented at the end of Study 6, underwent minor refinements and serves as the ultimate product of this action research, becoming the Final e-SEAT Framework.

Studies 4 to 6, including the four substudies of Study 4, make up six studies in total, and were a classic example of a product being evaluated through a series of action research studies. The research culminated in a final version that met user requirements. The final version, the ultimate product of this action research, is presented in Appendix J1. The interactive version of e-SEAT is also available on the UKZN server at: http://major.ist.ukzn.ac.za/upasana/SEATVer1/Home.aspx

The main outcome of Chapter 6 was that Research Questions 5 and 6 were answered. Question 5 asks 'What categories and criteria should be incorporated in a prototype framework to evaluate electronic assessment systems?'. The criteria identified through this research are the 182 in 11 categories which are presented in Tables 6.28 to 6.40. This framework is electronically administered and processed by the electronic e-SEAT Instrument. Selected e-SEAT screens are shown in Figures 6.8 to 6.14.

Research Question 6 asks 'How appropriate and effective is the proposed framework?' Through the evaluation, application and validation of the e-SEAT Framework in Studies 5 and 6, participants indicated e-SEAT it is appropriate for evaluating e-assessment tools and systems and that it does so effectively, thus giving a positive answer to Research Question 6. The reports from participants in Study 6 on their interaction with e-SEAT were even more positive than those in Study 5, specifically indicating that e-SEAT would be helpful in taking adoption decisions. Furthermore they appreciated the comprehensive set of evaluation criteria. The findings of both Study 5 and Study 6 provided data on the essential criteria for inclusion in the Evaluation Framework, as well as the appropriateness and effectiveness of the Evaluation Framework.

The ultimate deliverable of Chapter 6 is the e-SEAT Final Framework, presented in a concise form in Tables 6.30 to 6.42 and in full in Appendix J1.

Chapter 7, the conclusion of this study, is presented next.

CHAPTER 7 Conclusion and Recommendations

This chapter serves to summarise and conclude the research. Following an overall introduction in Section 7.1, the research questions that this study aimed to answer are revisited in Section 7.2 and the answers are concisely reviewed. Thereafter Section 7.3 presents the theoretical and practical contributions of this study. Section 7.4 overviews the methodological approach, which involves mixed methods research and a series of action research studies. Validity, reliability and triangulation, as implemented in this study, are discussed in Section 7.5. Finally, Section 7.6 addresses the limitations associated with this research, while recommendations for future research are presented in Section 7.7.

7.1 Introduction

The aim of this research was to develop, evaluate, refine, validate, and apply a framework for evaluating e-assessment systems being used, or under consideration for adoption, at higher-education institutions in South Africa. The envisaged contexts of use of the framework are schools and departments in Computing-related disciplines, namely: Computer Science (CS), Information Systems (IS) (also termed Informatics), and Information Technology (IT). The rationale for this work originated from a need identified in the South African context, as well as calls in the literature for increased research on the use of e-assessment and electronic testing systems.

The primary aim was to iteratively conduct research to support the development of the e-SEAT Framework (electronically Selecting and Evaluating an e-Assessment Tool) that can be applied by educators to facilitate the adoption of electronic assessment at their institutions and the selection of appropriate e-assessment systems/tools. e-SEAT can also be used to evaluate systems already in operation.

The secondary aim was to gain insight into the current extent, nature, and satisfaction of academics who were users of e-assessment tools. The study focused on local usage in South Africa. Although it was a secondary aim, this research was conducted at an early stage of the work, since its findings were used in the development of the Framework.

The series of studies involved:

- literature reviews, based mainly on international sources,
- data regarding local experiences with e-assessment in South Africa,
- quantitative and qualitative studies to
  - gather information on the current extent and nature of adoption and usage of e-assessment tools, as well as levels of satisfaction with such tools,
  - identify the types of e-assessment questions commonly adopted in e-assessment in South Africa,
  - understand the role of multiple choice questions (MCQs) in testing higher order thinking skills (HOTS),
  - develop the e-SEAT Framework to
    - assist users in the selection and acquisition of e-assessment tools, and
    - provide design guidelines for developers, and
  - evaluate, refine and apply the newly-developed evaluation framework through a series of action research studies.

The underlying research design of this work was action research (Sections 4.2.1 and 4.2.3). The action research process consisted of two phases. The studies in Phase 1 provided a background for generating SEAT and e-SEAT by investigating the local adoption of e-assessment systems and the types of questions these systems supported. After completion of Phase 1, Phase 2 focussed on the iterative generation of the initial evaluation framework named SEAT (Selecting and Evaluating an e-Assessment Tool). This involved creating, evaluating and refining SEAT. Evaluation criteria to populate the evaluation framework were compiled both from literature studies and from empirical findings of user-based surveys via custom-designed questionnaires and interviews. Thereafter, SEAT was converted to an electronic version, e-SEAT, which evolved through a similar iterative series of development, evaluation, refinement, application and validation via further questionnaire and interview research among participants from tertiary institutions in South Africa.

7.2 Research questions revisited

The main research question for this study was: "How does an academic evaluate an e-assessment tool, to identify the best-fit for his/her requirements?"

This question resulted in six subquestions, which are revisited in this section. Table 7.1 (similar to Table 1.1) presents the research questions and the chapters in which they were addressed in this study.

Table 7.1: Research questions and chapters in which they are answered

	Research question	Chapter(s) in which answered
RQ 1	What is the extent and nature of use of electronic assessment in Computing-related departments in South African universities?	5 (Section 5.1 and Section 5.2)
RQ 2	What types of questions are being adopted in e-assessment systems in South Africa?	5 (Section 5.3)
RQ 3	How appropriate are these questions (identified in Research Question 2) for testing higher order thinking skills (HOTS)?	5 (Section 5.3.3)
RQ 4	What are the requirements for selecting or personally developing an electronic assessment tool? • Theory: What does the literature suggest as appropriate requirements for electronic/online testing and assessment tools? • Practice: What criteria are used in practice in South African higher education institutions for the selection and use of electronic/online testing and assessment tools?	3 (Section 3.2.6; Table 3.1 Table 3.1) 5 (Section 5.4; Table 5.35)
RQ 5	What categories and criteria should be incorporated in a prototype framework to evaluate electronic assessment systems?	6 (Sections 6.1.1; 6.1.2; 6.1.3)
RQ 6	How appropriate and effective is the proposed framework?	6 (Sections 6.1.4; 6.2; 6.3)

Research Question 1 was answered in Studies 1 and 2. Study 1 specifically established the context of adoption of e-assessment tools within Computing-related academic departments/schools at South African tertiary institutions. Through the use of used open-ended questions in Study 1, the researcher was further able to investigate academics' satisfaction with the use of e-assessment tools. The 36 respondents in Study 1, representing eight universities, came from Computing disciplines – thirteen from Computer Science, eighteen from IS or IT and the rest from other Computing-related domains. They included both users of e-assessment and potential users. Study 2, which involved interview follow-ups to Study 1, targeted 72 users of e-assessment from 11 universities. In order to tap a larger base of regular adopters, respondents in this study were acquired from Computing and Non-Computing departments. The interviews addressed the same issues as Study 1, but in more depth.

Research Question 1	***What is the extent and nature of use of electronic assessment in Computing-related departments at South African universities?***

In Study 1, the extent of South African usage of e-assessment was found to be low, but on the increase. In fact, 38% of the users had adopted e-assessment only in the previous one or two years. Most of the users adopted mainly multiple choice questions and true/false. With regards to the nature of use, a variety of tools were being used by different users and at the various institutions. e-Assessment was employed more for formative than for summative assessment, thus indicating it was used for practice and revision as well as for formal testing. Usage was concentrated in first-level classes and large classes. In qualitative responses, the main reasons given for using e-assessment were the immediacy of feedback, consistent marking, and the ease of administering more assessments and frequent assessments. Among the disadvantages, participants acknowledged the complexity and time-consuming nature of creating good questions. It was also stated that technical issues and the need for administrative and technical support were barriers. A few mentioned that MCQs tend to test low levels of cognition. Benefits and disadvantages/barriers mentioned by respondents correspond well with those identified in the literature study.

The data obtained from the interviews conducted in Study 2, was similar. It confirmed, extended and elaborated the findings of Study 1. A high occurrence of new adopters was identified (53%), demonstrating that the adoption of e-assessment was on the increase at South African tertiary institutions. The use of a variety of tools was reported and they were different from the tools mentioned in Study 1. However, 74% of the participants were using tools embedded in the learning management system of their institution. This fact, as well as other open-ended responses, showed that the use of e-assessment was being increasingly supported at tertiary institutions, although most of the interviewees were free to use e-assessment as they wished. Again, high use was reported of formative assessment. For almost half of the participants, e-assessment contributed less than half of the final mark; and only 15% indicated that it contributed more than 80%. This evidences that it was used in a balanced manner with other means of assessment. As in Study 1, usage was concentrated at first-level and with large classes. Limited use occurred with postgraduate classes. As in Study 1, benefits and disadvantages associated with e-assessment were identified. The benefits outweighed the disadvantages, and the use of e-assessment contributed to productivity in terms of increased frequency of assessment, question reuse, and reduction of academics' workloads.

The findings of Study 2 also contributed to determining participants' requirements for e-assessment tools and identifying evaluation criteria for the selection and use of such tools. These criteria were incorporated in Table 5.35, which lists items obtained from interviews that are deemed necessary for inclusion in the SEAT Framework. Table 5.35 is part of the answer to Research Question 4, particularly its second subquestion. Research Question 4 is given after the report on Study 3.

Study 3 was a large study involving 92 respondents, of which the data of the 64 South Africans from fifteen institutions is reported here. Sixty percent (60%) were from Computing-related disciplines. Due to the multi-disciplinary nature of Computing disciplines, they are situated in various bases in different universities, namely Faculties of: Science, Engineering and Technology; Natural Sciences; Commerce and Economic Sciences, Management Sciences; and Higher Education (where e-learning is a sub-domain). Study 2 had identified that academics make particular use of basic MCQs and that only a few adopt them for more than just recall (lower cognitive levels). The questionnaire in Study 3 thus served a dual role, investigating the different types of MCQs adopted as well as their relevance to higher order thinking skills (HOTS).

Study 3 aimed to establish answers to Research Questions 2 and 3:

Research Question 2	***What types of questions are being adopted in e-assessment systems in South Africa?***
Research Question 3	***How appropriate are these questions (identified in Research Question 2) for testing higher order thinking skills (HOTS)?***

The investigation of the varying MCQ types adopted by participants, was in response to Research Question 2 and served to confirm the items identified in the literature. Sixteen types of questions of the multiple-choice genre were listed in the questionnaire and participants indicated all the types they were using or had used. The most common types were Multiple choice: single response and multiple response; followed by True/False (with or without explanation), Fill-in-the blank and Simulations (the latter being more common among Computing participants than Non-Computing). Matching items (including single matching and extended matching) were used more by Non-Computing participants than by Computing. There was use of Short answer, textual-input questions, but they require manual marking. The remaining types: Selection, Ranking, Diagrams/Video clips, Drag-and-drop, Reordering/sequencing, Categorising and Hotspots, had very low usage. Examples of all these types of questions are given in Section 3.1.5.

With regards to Research Question 3 on testing higher order thinking skills (HOTS), participants were asked to rate the sixteen types in terms of their relevance to HOTS. It became clear that the only type of question all participants were familiar with was multiple choice: single response. Nevertheless, significant numbers of responses rated the following types as being relevant to assessing HOTS:

Diagrams/Video clips, Simulation, Multiple choice: single response, Multiple choice: multiple response, Reordering/sequencing, and Categorising. Participants also rated the question types in terms of their suitability for the different levels of study. These responses rated the straight Multiple choice: single response, as being suitable for first-year assessment, gravitating through to Multiple choice: multiple response, Extended matching, True/False with explanation, and Simulations as being appropriate for fourth-level modules.

These responses identify a gap and evidence a vital need to orient and train academics in the creation and use of the less common question types and formats and, particularly, to encourage their use as an effective means of assessing HOTS and for assessment of exit-level modules.

As was the case with Study 2, the findings of Study 3 contributed to the compilation of criteria for Table 5.35, which in turn contributes to the SEAT Framework. Study 3 thus provided part of the answer to Research Question 4. In particular, it drew attention to the need for a category on Question Types to be included in the SEAT Framework (Table 5.35).

The initial generation of the SEAT Framework is now addressed. Criteria that are appropriate for inclusion in a framework for evaluation of e-assessment systems, were identified from the literature (Table 3.1 – which answered the first part of Research Question 4), as well as empirically in Studies 2 and 3 (Table 5.35 – which answered the second part of Research Question 4). These tables provide a comprehensive list of criteria that answered Research Question 4 and that served as the foundation of the SEAT Framework.

Research Question 4	***What are the requirements for selecting or personally developing an electronic assessment tool?*** • ***Theory: What does the literature suggest as appropriate requirements for electronic/online testing and assessment tools?*** • ***Practice: What criteria are used in practice in South African higher education institutions for the selection and use of electronic/online testing and assessment tools?***

After Studies 1, 2 and 3 (Phase 1 of the action research), the criteria identified in the Literature Study in Chapter 3 and presented in Table 3.1, and the criteria used in practice, originating from the empirical work in Chapter 5 and culminating in Table 5.35, were integrated to synthesise the first version of the prototype SEAT Evaluation Framework. The research then moved on to Phase 2 comprising Studies 4, 5 and 6, which involved research activities such as evaluations, validations and applications of SEAT and e-SEAT. In line with action research, the researcher reflected and responded to the outcomes of each study in the action research series, and planned the following and next few studies, that ultimately culminated in the e-SEAT Final Framework.

In Study 4 the initial version of the comprehensive framework was iteratively evaluated, applied and refined in four substudies, each of which moved SEAT a step closer to the final SEAT Framework. The first three of these substudies, Studies 4a, 4b and 4c, answered Research Question 5.

Research Question 5	***What categories and criteria should be incorporated in a prototype framework to evaluate electronic assessment systems?***

The initial categories of criteria were re-structured by the researcher into ten categories. The framework was then hosted on the online Survey Monkey tool to facilitate access for the participants in the next studies, which would result in iterative refinement of SEAT.

The Pilot Study, Study 4a, was used to obtain essential feedback from two participants on the design, content and validity of the first-draft version of the framework. It also served to try out the research approaches prior to Study 4b. The participants in the Pilot rated the criteria according to Likert scaling to indicate their perceived relevance and appropriateness. In addition, they provided valuable and detailed general comments that resulted in rewording of criteria, correction of ambiguous terminology, migration between categories, and merging/removal to avoid duplication. They also suggested further categories and criteria. Table 6.3 lists their feedback and suggestions, and indicates the researcher's responses and resulting adjustments to the SEAT Framework.

The Evaluation Study, Study 4b, was an extensive study with 56 participants from 16 universities. The aim was to determine which of the criteria in this version of SEAT were essential for any e-assessment tool. Participants rated each criterion on a scale of 1 to 7 to indicate how they perceived its importance. The mean ratings were statistically analysed to assist the researcher in establishing which criteria were essential and which could be considered for removal. The questionnaire also included open-ended questions for general comments and concerns. This provided some qualitative data. Table 6.17 lists the participants' feedback and suggestions, and indicates the researcher's responses and resulting adjustments to the SEAT Framework.

In the process of providing category-by-category results of the ratings and statistical analyses, ten tables in Section 6.1.2 respectively list the ten categories of the evaluation framework and the criteria they contained at that stage. These tables thus play a second important role by providing the first complete presentation of the categories and criteria in an early version of the SEAT Framework.

After Studies 4a and 4b, the resulting version of SEAT reached the stage of an operational, though manually operated, prototype. Study 4c was a Proof of Concept Study. Its participants were three South African experts in the fields of e-assessment and MCQs, from three different universities. They each conducted a rigorous, in-depth overview of SEAT and were interviewed by the researcher. Their contributions and insights added value and led, among others, to structural changes and semantic improvements. The most important were that an additional (eleventh) category was included and, secondly, the categories were classified into Functional and Non- Functional, leading to five categories of Functional Criteria and six categories of Non- Functional criteria. Table 6.22 lists the experts' feedback and suggestions, and indicates the researcher's responses and resulting adjustments to the SEAT Framework.

Studies 4a, 4b, and 4c thus investigated the categories and criteria within SEAT and contributed to their evolution.

The final research question, Research Question 6, considered the appropriateness and effectiveness of the Evaluation Framework. Three studies, Study 4d, Study 5 and Study 6, each contributed to answering this question.

Research Question 6	*How appropriate and effective is the proposed framework?*

The Application Study, Study 4d, provided part of the answer to Research Question 6, by requiring participants to apply SEAT to evaluate an existing e-assessment system. The participants were seven South Africans experts from five different universities, five of whom had used e-assessment extensively for over five years. Four participants provided qualitative comments, stating that the Framework was a valuable instrument and appropriate for its purpose. They would like it to be deployed in their institutions. No shortcomings were identified.

The SEAT Framework was developed over the four iterations of Study 4. Findings of the substudies were used to improve the framework as it evolved from the Pilot Framework in Study 4a, to the Evaluation Framework in Study 4b, to the Proof of Concept Framework in Study 4c, and finally to the Application Framework in Study 4d. These substudies were conducted on the manually operated SEAT Framework, hosted on Survey Monkey.

At this point SEAT was converted to an interactive electronic version, e-SEAT (electronically Selecting and Evaluating an e-Assessment Tool), with an automated scoring and calculation system and a reporting facility. A computer programmer did the coding to implement the researcher's design and create the first functional prototype of the e-SEAT instrument. The electronic instrument became a platform on which to implement and deliver the SEAT Evaluation Framework. Studies 5 and 6 were designed to complete the answer to Research Question 6 regarding the appropriateness and effectiveness of the proposed framework (e-SEAT).

In Study 5, the e-SEAT Evaluation, participants evaluated the e-SEAT Framework. The four participants from four universities were expert users of e-assessment, who had not taken part in any of the Study 4 substudies, which gave them unbiased perspectives. They accessed e-SEAT via the UKZN server and used the instrument to evaluate an e-assessment system they were currently using, after which they completed an evaluation questionnaire. Nine themes emerged from content analysis of their interviews and qualitative responses to the questionnaire. The findings were positive; a few shortcomings were mentioned, but very few changes were suggested. Some additional processing features were recommended. Table 6.26 lists the experts' feedback and suggestions, and indicates the researcher's responses and resulting adaptations to the e-SEAT Framework. A valuable suggestion was that the concluding report should be e-mailed to the user. This feature was incorporated by the programmer.

Study 6, the e-SEAT Application and Validation Study, was conducted by three specialist e-assessment users from three universities, who critically reviewed the framework and stringently validated it. A secondary aim of Study 6 was the application, by each participant, of the e-SEAT Validation Framework to an e-assessment tool they used. Their reports on interaction with e-SEAT were even more positive than those in Study 5. They experienced e-SEAT as useful and intuitive to use, and found the reports that it generated would be helpful in taking adoption decisions. They appreciated the comprehensive set of evaluation criteria. One participant raised an issue that that been encountered in Study 5, namely that disorientation could occur within the framework and a progress bar was needed. Table 6.29 lists the specialists' feedback and suggestions, and indicates the researcher's responses and resulting adjustments to the e-SEAT Framework.

The evaluation, application and validation of the e-SEAT Framework in Studies 5 and 6, indicate that it is appropriate for evaluating e-assessment tools and systems and that it does so effectively, thus giving a positive answer to Research Question 6.

Through action research processes, e-SEAT evolved from the e-SEAT Evaluation Framework in Study 5, through the e-SEAT Validation Framework in Study 6, to the ultimate product, the e-SEAT Final Framework, which is a basic interactive instrument, on which to implement the SEAT Framework. This prototype e-SEAT does not have full functionality and, at this stage, should not be viewed as a system for general dissemination, but further development and future use is anticipated. Section 7.4 on the methodological approach of this research – mixed methods and action research – highlights the role and value of action research as the underlying research design of this work.

Main Research Question

"How does an academic evaluate an e-assessment tool, to identify the best-fit for his/her requirements?"

Through the six studies in the Action Research series, this research led to the development and evolution of a framework that facilitates the evaluation of e-assessment systems/tools. Each study in the series conducted in this research, played a pivotal role in the evolution of the SEAT Framework and the e-SEAT Instrument. Furthermore, they contributed to answering the six subquestions and thus culminated in answering the Main Research Question.

Figure 7.1 graphically depicts the role of each of the six studies in answering the research questions.

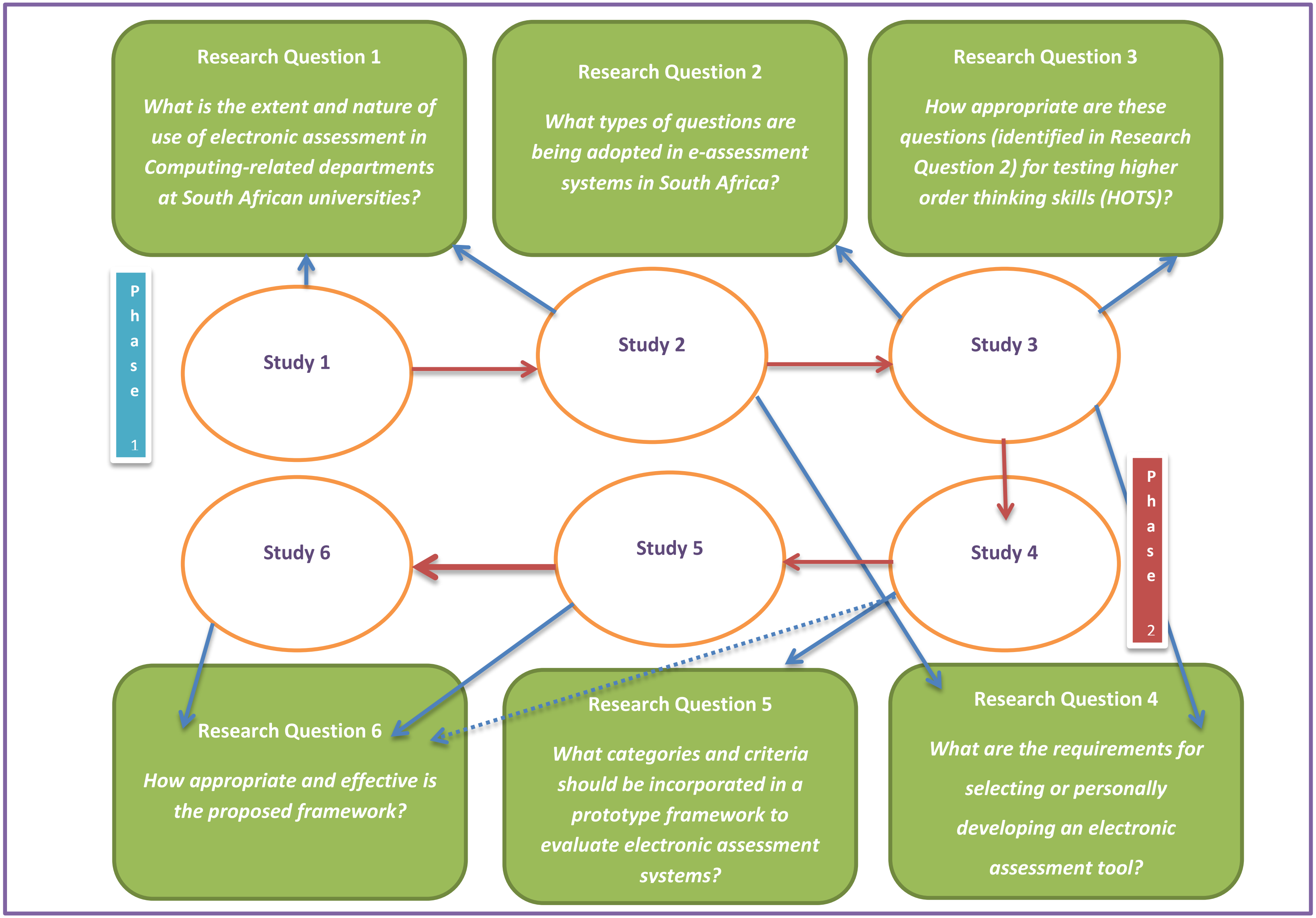


Figure 7.1: Relationships between studies conducted and research questions posed

7.3 Practical and theoretical contributions of this study

This research identified a niche area by acknowledging the need for an evaluation framework to evaluate e-assessment systems of the MCQ genre. It made a major contribution to the body of knowledge on e-assessment by presenting an extensive list of evaluation categories and criteria, which were incorporated in the SEAT Framework to contribute to a conceptual understanding of requirements and features for e-assessment systems that administer MCQs. The Framework was then converted to e-SEAT, an interactive version that automates processing and presentation of the results. These contributions and others are briefly presented below.

7.3.1 Practical contribution of the study: e-SEAT

The primary practical contribution of this research is the innovative interactive e-SEAT Instrument for evaluating e-assessment systems and tools. e-SEAT is the interactive medium on which the theoretical SEAT Framework resides and is delivered to users. It can assist users and potential users of e-assessment in examining e-assessment systems they are considering adopting. The automated results returned by e-SEAT give scores per category and totals that assist users in selecting a system/tool and in comparing systems to determine which would be most appropriate to meet their requirements. e-SEAT can also be employed for evaluating systems in use to establish their fitness for purpose.

A further practical contribution of this work, is the presentation in Sections 3.1.4 and 3.1.5 of the varying types of questions of the MCQ genre, including the less common types. Each question type is illustrated by an authentic example from a Computing-related module.

7.3.2 Theoretical contribution of the study: SEAT Framework

The researcher has contributed to the body of knowledge on e-assessment by compiling the SEAT Framework (Appendix J1) as a comprehensive and well-structured set of categories and criteria relating to features of e-assessment systems. Table 3.1 indicates how the initial set of categories and criteria was derived from the literature. Other literature sources discussed in Chapter 3 further affirmed these criteria, and the empirical studies reported in Chapter 5 and 6 resulted in the evolution of the Framework.

SEAT deepens understanding of the requirements for, and implementation of, e-assessment tools. It can also serve as a comprehensive set of design guidelines from which designers and developers of e-assessment can select guidelines and customise them for systems under production.

Figure 7.2 shows the main components and evaluation categories of the e-SEAT Instrument (Appendix J1) that was generated by the conversion of the conceptual SEAT Framework to an electronic instrument.

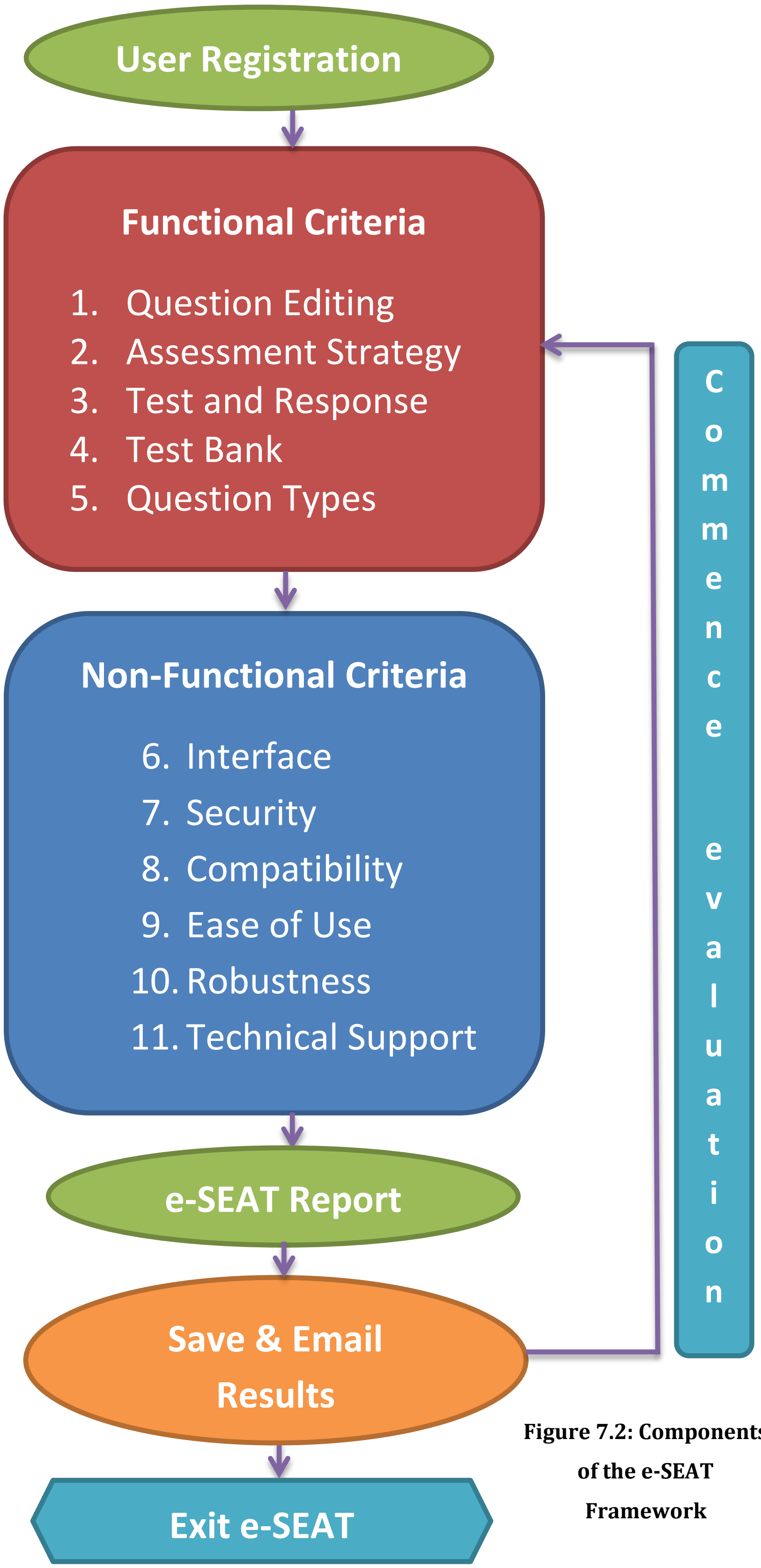


Figure 7.2: Components of the e-SEAT Framework

7.4 Mixed methods research and action research

7.4.1 Mixed methods research

This research (see Section 4.1.4) adopted a mixed methods approach (Section 4.1.3.3); combining quantitative (Section 4.1.3.2) and qualitative studies (Section 4.1.3.1). Quantitative research collects numerical data which is typically analysed using statistics, while qualitative studies collect and analyse verbal or written data.

Combining qualitative and quantitative data had a synergistic impact, as the mixed method dual strategy produced both confirmatory and complementary findings, where one method produced data that was not possible with the other. In this study surveys – questionnaires and interviews – were conducted to obtain quantitative and qualitative data. Real-world activities were also undertaken in some studies. Questionnaires were used in Study 1 and in Study 3 (a large sample, 92) to gather numerical data without the researcher being physically present. The questions were structured, closed and numerical, making the responses simpler to analyse. After the quantitative work in Study 1, in-depth qualitative data was collected by follow-up interviews in Study 2, where 68 interviews were conducted over four months. These interview responses complemented and elaborated information that had emerged from Study 1. For example, the benefits and disadvantages of e-assessment given by participants in Study 1 were confirmed in the interviews in Study 2.

In Study 4 which focused on refining and consolidating the criteria in the SEAT Framework, a variety of complementary methods were employed to develop richer insights into the area of interest, in line with mixed methods. The mainly quantitative Evaluation in Study 4b involved 56 participants. Its processes and terminology were tried out in the small qualitative Pilot Study, Study 4a which pilot-tested the strategies and questions, except that Study 4b was delivered on Survey Monkey, while Study 4a used a paper-based version of SEAT and included personal contact between the researcher and the two participants. The Pilot served its purpose; it facilitated Study 4b and also ensured that the rating options built into the SEAT Framework on Survey Monkey catered for all possible responses. One of the main purposes of Study 4b was to statistically identify low-rated criteria that should be considered by the researcher as candidates for possible removal. Study 4c (Proof of Concept) and Study 4d (Application of SEAT) with small samples of experts were highly qualitative.

They involved in-depth interviews, open-ended topics and manual discourse analysis. The Study 4c data included rich insights and resulted in recommendations for far-reaching refinements to the SEAT Framework. Most of the qualitative data gathered in Study 4c was spontaneous and unprompted, yet it confirmed statements made by participants responding to specific questions in other studies. The participants in Study 4d undertook a real-world activity. They investigated SEAT in operation, using it effectively to evaluate authentic e-assessment tools.

Study 5 was a small-scale, in-depth evaluation of the interactive electronic version, e-SEAT. It produced some quantitative, but mainly qualitative, data. Nine themes emerged and were used to structure presentation of the findings. Study 6 was purely qualitative, where three e-assessment specialists separately and critically reviewed e-SEAT to validate it. They each used it successfully to evaluate an e-assessment system they had used previously.

7.4.2 Action research

Action research, which is described in Sections 4.2.1 and 4.2.3 focuses on practical change and generates solutions to real-world problems. It is based on an iterative, researcher-centric series of cycles, with each cycle involving planning, actions, observing, reflecting and responding. The research should result in a final product or practical intervention that assists in solving the problem.

The present research originated with a practical problem, namely the challenge faced by adopters of e-assessment who need support in the selection and implementation of such a system or tool. Action research is participative, which was the case here. After networking to find stakeholders, the researcher engaged personally with many of them. Phase 1 of the action research involved using Studies 1, 2 and 3 to cumulatively build a foundation of knowledge on the extent and nature of e-assessment adoption by South African Computing academics. This set a context for research on requirements for, and evaluation of, e-assessment systems as a background to developing the SEAT Framework. Knowledge was also obtained on use of the less common question types, and this informed the need to include question types as a category in the evaluation framework.

Criteria for evaluating e-assessment systems that administer questions of the MCQ genre, were identified from the literature and from empirical studies. They were classified into categories that were input to Phase 2 of the action research process that created and refined, first, the SEAT

Framework of categories and criteria and then the interactive electronic e-SEAT instrument for evaluating and scoring e-assessment systems. Using the cyclic action research processes of planning, actioning, reflecting and responding, the researcher conducted the empirical Studies 4a, 4b, 4c, 4d, 5 and 6, which evaluated, applied, refined and validated the evolving Frameworks through the versions in Figure 6.2. The studies are described in detail in Section 7.2, which provides answers to the Research Questions.

The action research approach served well for this research and added value with its iterative approach of successive studies over a period of four years. Each study was separate and had a differing purpose and a different perspective. The feedback of each study was consolidated by the researcher into a table listing recommended changes to the relevant Framework, such as rewording of criteria, addition/removal of criteria, elimination of duplicates, merging, and suggestions of new features. Studies 4a, 4b, 4c, 5 and 6 culminated in Tables 6.3, 6.15, 6.20, 6.24 and 6.27 respectively. These tables show the evolution of the two products, the SEAT Framework and the e-SEAT Framework, as they went through successive refinements. The number of recommended changes decreased over the series, as the frameworks became increasingly fit for purpose. There were over 100 participants in total and, in each table; the code names (e.g. PP5) were given of the participants who had made suggestions.

Table 7.2 commences by depicting how the SEAT Framework originated with the merger of criteria from the literature with criteria from empirical work. It demonstrates the evolution of SEAT into e-SEAT and the e-SEAT development process. It shows the changes in each category of criteria that resulted from the successive studies in the action research series.

PLEASE OPEN THIS PAGE TO VIEW A LARGE SIZE PRINTOUT OF TABLE 7.2, WHICH SUMMARISES THE EVOLUTION OF THE CRITERIA AND CATEGORIES IN THE e-SEAT FRAMEWORK

Table 7.2 Overview of e-SEAT Framework criteria evolution

ACTION SERIES	Literature Table 3.1	Interviews and Questionnaire 2 (Empirical Work) Table 5.35	Post Combination of Tables 3.1 and 5.35		Pilot Framework (Study 4a) Appendix F1	Post Pilot (Word Version)		Pilot Framework (Study 4a) Appendix F2	Post Pilot Study		Evaluation Framework (Study 4b) Appendix G	Post Evaluation Study	Proof of Concept Framework (Study 4c) Appendix H1	Post Proof of Concept Study		Application Framework (Study 4d) Appendix I	Post Application Study	Final e-SEAT (Appendix J1)
CATEGORIES			Inserted	Removed		Inserted	Removed		Inserted	Removed		Inserted		Inserted	Removed		Inserted	
Interface Design	4	4	1		9			9	1		10	4	14			14		14
Question Editing	19	5	2		26			26			26	7	33		2	31		31
Assessment Strategy	4	2	2		8			8	1		9	6	15		4	11		11
Test and Response Analysis	24	5	7		36	6		42	2		44	4	48	2	14	36		36
Test Bank	2	0			2			2			2	2	4	2	2	4		4
Security	10	4	1		15	3		18			18	5	23		2	21		21
Compatibility	4	2			6	2		8			8	1	9		1	8		8
Import/Export	1	0	1		2		2											
Ease of Use	11	2	1	1	13	4		17	7	1	23	2	25		2	23		23
Technical Support	11	2			13	3		16		6	10		10		1	9	2	11
Training	1	0			1		1											
Robustness														5		5		5
Question Types		16			16			16			16	1	17	1		18		18
Total Categories	11	12			12			10			10		10			11		11
Total Criteria	91	42			147			162			166		198			180		182
Total Criteria Inserted/Removed			15	1		18	3		11	7		32		10	28		2	

7.5 Validity, reliability and triangulation

This section considers certain theoretical concepts of validity, reliability and triangulation, initially introduced in Section 4.6, and indicates how they were implemented in the present research. In short, validity relates to accuracy of the findings, while reliability relates to consistency of the research approach, and triangulation involves strengthening the research by using more than one data sources or more than one perspective (Creswell, 2009).

7.5.1 Validity

Section 4.6.1 introduced theoretical concepts of validity, as outlined by various authors (Cohen et al., 2011; Creswell, 2009; Oates, 2010). Certain of these concepts are revisited in Table 7.3, showing how they were implemented in this research.

Table 7.3: Validity in this research

Concept	How implemented
Accuracy of the findings	
Adoption of an appropriate process	Action research with its longitudinal, participative process is highly appropriate to refining products (Section 7.4)
Findings can be linked back to the data	Qualitative and quantitative findings were derived from actual responses of participants as data. Qualitative findings were extracted and reported under common themes (Sections 5.1.6, 5.2.4, 6.1.1, 6.1.3, 6.1.4, 6.1.5, 6.2.1 and 6.3.1). Quantitative findings were presented as summaries of various statistical analysis techniques (Sections 5.1.5, 5.2.3, 5.3 and 6.1.2)
Findings answer the research questions	See Section 7.2 – each research question has a directly related answer.
Validity of various types of research **Qualitative data validity, addressed by honesty, depth and richness of data**	Honesty was demonstrated by responses that were critical of both SEAT and e-SEAT. This was displayed by participants who provided both positive and negative comments when reviewing the frameworks (For example, see Sections 6.1.3.2. 6.3.1.5)

	Care was taken in developing instrumentation to facilitate rich data collection. For example, rating and open-ended questions were provided in all questionnaires, as well as in the interviews, to allow participants to express their views clearly, and clarify any ratings they selected in the quantitative based questions.
Quantitative data validity	Analysis: use of more than one statistical method on the same data, for example in Section 6.1.2.
Mixed-methods validity (legitimation)	
– **Representation**	Some of the data used mainly words to capture experiences, for example, verbatim extracts from the interviews in Study 2 (Section 5.2.3), the Proof of Concept study (Section 6.1.3), and follow up interviews to clarify experiences with SEAT (Sections 6.2.1.5 – 6.2.1.10)
– **Results that are dependable, credible, transferable, plausible, confirmable and trustworthy**	Participants with expertise and integrity were carefully selected: Academics (Table 5.1), users of e-assessment (Table 5.29 and text following) and experts/specialists in the field (Tables 6.24, 6.27 and text after it).
– **Integration**	Quantitative and qualitative methods were integrated in the action research studies. Some studies were quantitative, some were qualitative, and some had both quantitative and qualitative aspects. (Section 5.1.2, the part about Research Question 1)
Non-representativeness – threats **Too few participants**	Not a threat. Total number of participants in the six studies was 108. Some participants took part in more than one study, so number of distinct participants was 84.
Over-reliance on specific types of participants or non-representative participants.	Not a threat. Participants were from a range of ranks: (e.g. junior lecturer to professor. Besides academics, there were participants of other occupations, e.g. senior managers, e-learning developers.) Participants ranged over four main racial groups in SA; ages between 30 and 58; 16 institutions were represented over the 9 action research studies

Content validity **Interviews** **Surveys**	Verbatim quotes in Sections 5.2.4.1 and 6.2.1 Pilot testing in Study 4a to ensure that items in instrument of Study 4b were concise, relevant, and unambiguous.
Member checking	Follow up interviews were conducted to clarify comments made by participants in the Proof of Concept study (Section 6.1.3) Study 5 included interviews (Table 6.24 and text afterwards). Transcripts were sent to participants for verification prior to analysis of the data.
Clarifying bias	Application of the researcher's knowledge in reviewing criteria identified for possible removal from the SEAT Framework, subsequent to the statistical analysis (Section 6.1.2.4)
Presenting negative information	Both positive and negative comments are reported (Sections 6.1.1.1 - 6.1.1.6, 6.2.1.6, and 6.3.1.5)
Prolonged time in the field	The researcher conducted personal interviews in Study 2 (Section 5.2.2) and took the opportunity for informal observation of the tools in use.

7.5.2 Reliability

Section 4.6.2 introduced theoretical concepts of reliability, as outlined in the literature (Cohen et al., 2011; Oates, 2010). Some of these concepts are revisited in Table 7.4, showing how they were implemented in this research.

Table 7.4: Reliability in this research

Concept	How implemented
Consistency and trustworthiness	Repeating questions from Questionnaire 1 in Study 1 during the interviews held in Study 2
Credibility	Careful selection of participants in Studies 4 to 6, in particular the small and select samples of e-assessment experts and specialists in Studies 4c, 4d, 5 and 6
Neutrality	No personal biases of participants or the researcher were identified in the study
Dependability	Interview transcripts were validated by participants prior to inclusion in the findings
Internal consistency	Several statistical analysis techniques were adopted
Repeatability	The researcher observed different participants doing the same task, thus confirming that each participant understood what was expected and undertook the required task in a similar manner.

7.5.3 Triangulation

Section 4.6.3 introduced the theoretical concept of triangulation, as outlined by various authors (Cohen et al., 2011; Creswell, 2009; Oates, 2010). Certain of these concepts are revisited in Table 7.5, showing how they were implemented in this research.

Table 7.5: Triangulation in this research

Concept	How implemented
Methodological triangulation	This research adopted method triangulation, making use of surveys, interviews and informal observation for data collection.
Strategy triangulation	Qualitative and quantitative strategies were combined as parts of a mixed methods approach
Time triangulation	Longitudinal action research studies done over a 4-year period
Space triangulation	International participants were part of this study (through the use of a Computing-related mailing list); In all 9 studies, participants represented various cultures of South Africa
Investigator triangulation	N/A to PhD research, which may not be collaborative work
Theoretical triangulation	N/A

7.6 Limitations of the research

This research is focused on the practices and perspectives of academics in tertiary institutions who are adopters of e-assessment. It does not investigate students' viewpoints on e-assessment.

This research is situated in the subdiscipline of multiple choice question (MCQ)-related assessment. Thus, the context of the evaluation framework is restricted to questions from the MCQ genre, including the less commonly-used forms such as fill-in-the-blank, true/false with explanation, matching columns, extended matching, hotspots on diagrams, drag and drop, categorising, and simulations.

The research is aimed at supporting South African academics in Computing-relating disciplines in adopting and using e-assessment, and was conducted mainly within higher educational institutions in South Africa.

A basic interactive electronic instrument, named e-SEAT, has been developed, on which to implement the SEAT Framework. This prototype e-SEAT has limited functionality and should not be viewed as an operational system for public use.

The immediate context of this research is restricted to the application of e-assessment in controlled testing environments.

As stated in the Disclaimer in Section 1.1, the use of artificial intelligence techniques for analysing textual responses is excluded from the present research. Similarly, the study does not investigate the use of text analysis tools, such as those implemented by sophisticated pattern matching techniques, and natural language processing.

The framework developed as a result of this research is an evaluation framework and not a conceptual framework.

This research does not investigate other forms of e-assessment, such as e-portfolios, blogs, wikis, or peer-assessment.

7.7 Recommendations and future research

7.7.1 Recommendations

The responses regarding the use of different kinds of MCQs identify a gap and evidence a need to orient and train academics in the creation and use of the less common question types and formats discussed in Sections 3.1.4 and 3.1.5. In particular, academics should be encouraged to use them as an effective means of assessing HOTS and for assessment of exit-level modules.

The basic interactive electronic instrument, e-SEAT, should be converted from a prototype with limited functionality to an operational system for public use.

The development and sharing of questions that can be used in the same subject area, irrespective of the textbook prescribed, and storing them in a common database accessible to academics across higher educational institutions in South Africa, would be very valuable.

7.7.2 Future research

The criteria that comprise e-SEAT also provide a comprehensive set of design guidelines for designers and developers of e-assessment systems. Future research could involve a design research venture of building and evaluating an e-assessment system based on the criteria within SEAT.

The official extension of this research to international tertiary institutions as a collaborative venture, would add a global perspective to the local domain of e-assessment.

Research should be conducted to ascertain students' perceptions of the adoption of e-assessment.

Further research of the same nature, should be conducted among Computing academics in higher educational institutions in South Africa who are adopters of e-assessment, to establish the situation five years after this work commenced.

This research could be expanded to include other forms of e-assessment, such as, e-portfolios, blogs, wikis, peer-assessment.

Investigation of international trends on e-assessment adoption would prove useful.

The link between the e-assessment tool adopted and the LMS in use at the Institution could be elaborated in future research, specifically addressing if the assessment:

- could be embedded in the learning modules, and
- can inform a students' pathway through the LMS.

7.8 Chapter summary and conclusion

This chapter presented a summary of the work undertaken. Following a brief introduction, the six research questions that this study aimed to answer, were revisited and answered one by one, thus briefly summarizing the studies conducted and their findings. Thereafter the practical and theoretical contributions of this study were presented. The primary practical contribution is the interactive electronic e-SEAT instrument for evaluating e-assessment systems and tools. The main theoretical contribution is the compilation of the SEAT Framework as a comprehensive and well-structured set of categories and criteria relating to features of e-assessment.

The methodological approaches to this series of nine studies were overviewed. The underlying research design was action research. Its longitudinal and participative nature showed it to be suitable for addressing the research problem, namely the need for an instrument to support academics in evaluating and selecting appropriate e-assessment tools. The work was done in two phases. Phase 1 comprising three studies, laid a theoretical foundation by investigating the adoption of e-assessment in South African Computing-related disciplines, while Phase 2 employed three further studies to iteratively create and refine the SEAT Evaluation Framework. Each study and substudy closed with a table of adaptations to the Framework, which resulted from the research conducted in that study. The other methodological strategy was the mixed methods approach, which combined qualitative and quantitative studies in a synergistic way that used them to complement and confirm the findings.

Validity, reliability and triangulation, as achieved in this research study, were discussed. Finally the limitations associated with this study, the resulting recommendations, and directions for future research, were presented.

To reiterate, Chapter 1 of this study presented the introduction and background of this research. The literature studies in Chapters 2 and 3 outlined the terms, concepts, and attributes associated with assessment in general and e-assessment in particular. They reported on international usage of and practices in e-assessment, and also outlined an initial synthesis of evaluation criteria from which the evaluation framework for this study was derived. Chapter 4 set out the research design and methodology adopted in this study. Chapters 5 and 6 discussed the results from the data collection of Study 1 to Study 6. Chapter 5 focused on Phase 1 of the action research process. Phase 1 included Study 1 which identified the extent and nature of use of assessment tools in the context of South African computing education and Study 2 which used interviews to expand Study 1. Study 3 presented the investigation into the types of MCQs adopted by South African academics, as well as the applicability of these to higher order thinking skills (HOTS). The development of the SEAT Framework was facilitated by Study 4 with its four substudies, ranging from the Pilot Study (Study 4a), through to the Evaluation Study (4b), Proof of Concept Study (4c) and an Application Study (4d), was discussed in Chapter 6. These four substudies related to the evolution of the SEAT Evaluation Framework of categories and criteria. The further work done in Phase 2 of the action research series involved Study 5 and Study 6 which related to the creation and refinement of the prototype electronic version of the final e-SEAT Framework for evaluating e-assessment systems. This chapter, Chapter 7, provided a summary, recommendations and directions for future research. The List of References follows. The Appendices are included on a CD, attached to the back cover of this Thesis..

REFERENCES

[1] Akinsanmi, O., Agbaji, O. R. T. & Soroyewun, M.B. (2010). Development of e-assessment platform for Nigerian universities. *Research Journal of Applied Sciences, Engineering and Technology*. 2(2): 170-175.

[2] Alessi, S.M. & Trollip, S.R. (2001). *Multimedia for learning: methods and developments*. 3rd edition. Massachusetts: Allyn & Bacon.

[3] Alton, R. (2009). *How to write Multiple Choice Questions (MCQs)*, RLO-CETL - the Centre for Excellence in Teaching & Learning in Reusable Learning Objects. Available from http://rlo-cetl.blogspot.com/2009/05/how-to-write-multiple-choice-questions.html [Accessed: 4 January 2014].

[4] Amiel, T. & Reeves, T.C. (2008). Design-Based Research and Educational Technology: Rethinking Technology and the Research Agenda. *Educational Technology and Society.* 11(4): 29-40.

[5] Anderson, S., Ball, S. & Murphy, R. (1975). *Encyclopedia of educational evaluation*. San Francisco: Jossey-Bass.

[6] Anderson, T. & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher* 41(1): 16-25.

[7] Ardito, C., Costabile, M.F., de Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T. & Rossano, V. (2006). An Approach to Usability Evaluation of E-Learning Applications. *Universal Access in the Information Society.* 4(3):270-283.

[8] Ashcroft, K. & Palacio, D. (1996). *Researching into Assessment and Evaluation in Colleges and Universities. Practical Research Series*. London: Kogan Page.

[9] Bani-Ahmad, S. & Audeh, A. (2010). Re-engineering of multiple-choice exam from production tools: Cost-effective and quality-assurance approach. *Paper presented at the 2010 International Conference on Educational and Information Technology (lCEIT)*, 17-19 September 2010. China: Chongqing.

[10] Banks, S.R. (2012). *Classroom Assessment: Issues and Practices.* 2^{nd} Edition. Long Grove, Illinois: Waveland Press.

[11] Barab, S. & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences.* 13(1): 1-14.

[12] Baskerville, R. & Myers, M.D. (2004). Special Issue of Action Research in Information Systems – Making IS Research Relevant to Practice – Foreword. *MIS Quarterly.* 28: 329-335. September 2004.

[13] Belton, M. & Kleeman, J. (2001). New developments in perception designed to meet the changing face of CAA. *Proceedings of the 5th CAA Conference, Loughborough*. Loughborough University.

[14] Besterfield-Sacre, M. & Shuman, L. J. (2008). The Future of Assessment. In: *Designing Better Engineering Education Through Assessment: A Practical Resource for Faculty and Department Chairs on using Assessment and ABET Criteria to Improve Student Learning*. Spurlin, J., Rajala S. & Lavelle, J. (eds). Sterling. Va: Stylus Publishing.

[15] Beullens, J., van Damme, B., Jaspaert, H. & Janssen, P.J. (2002). Are extended-matching multiple-choice items appropriate for a final test in medical education? *Medical Educator.* 24(4): 390-395.

[16] Bialo, E. R. & Sivin-Kachala, J. (2013). IESD White Paper: *Promising Practices in the Use of Mobile Technology with eInstruction's Insight 360*.

[17] Biggs, J. (2003). Aligning Teaching and Learning Course Objectives. *Teaching and Learning in Higher Education: New Trends and Innovations*. April, 2003.

[18] Bleske-Rechek, A., Zeug, N. & Webb, R.M. (2007). Discrepant performance on multiple-choice and short answer assessments and the relation of performance to general scholastic aptitude. *Assessment & Evaluation in Higher Education.* 32(2): 89-105.

[19] Bloom, B. S. (1956). *Taxonomy of educational objectives, Handbook I: The cognitive domain*. New York, NY: David McKay Inc.

[20] Boyle, A. & Hutchison, D. (2009). Sophisticated tasks in e-assessment: what are they and what are their benefits? *Assessment & Evaluation in Higher Education.* 34(3): 305-319.

[21] Brink, R. & Lautenbach, G. (2012). Electronic assessment in higher education. *Educational Studies*. 37(5):503-512.

[22] Brown, S., Bull, J. & Race, P. (2013). *Computer-assisted assessment of students*. London: Routledge.

[23] Buchan, J.F. & Swann, M. (2007). A Bridge too Far or a Bridge to the Future? A case study in e-assessment at Charles Stuart University. *Australasian Journal of Educational Technology.* 23(3): 408.

[24] Bull, J. (1993). Using technology to assess student learning. In: *Committee of Vice Chancellors and Principals Staff*. Universities' Development Unit. United Kingdom.

[25] Bull, J. & McKenna, C. (2003). *A blueprint for computer-assisted assessment*. London: Routledge.

[26] Burton, R.F. (2001). Quantifying the effects of chance in multiple choice and true/false tests: question selection and guessing of answers. *Assessment & Evaluation in Higher Education.* 26(1): 41-50.

[27] Byrnes, R. & Ellis, A. (2006). The prevalence and characteristics of e-assessment in Australian universities. *Australasian Journal of Educational Technology.* 22(1): 104.

[28] Carter, J., Ala-Mutka, K., Fuller, U., Dick, M., English, J., Fone, W. & Sheard, J. (2003). How shall we assess this? *ACM SIGCSE Bulletin*. 35(4):107-123.

[29] Case, S.M. & Swanson, D.B. (1993). Extended-matching items: a practical alternative to free-response questions. *Teaching and learning in medicine.* 5(2): 107-115.

[30] Case, S.M., Swanson, D.B. & Ripkey, D.R. (1994). Comparison of items in five-option and extended-matching formats for assessment of diagnostic skills. *Academic Medicine.* 69(10): 51.

[31] Catley, P. (2004). One Lecturer's Experience of Blending E-learning with Traditional Teaching or How to Improve Retention and Progression by Engaging Students. *Brookes eJournal of Learning and Teaching.* 1(2): 1-8.

[32] Cheng, A.C., Jordan, M.E. & Schallert, D.L. (2013). *Reconsidering assessment in online/hybrid courses: Knowing versus learning.* Computers & Education.

[33] Christakoudis, C., Androulakis, G.S. & Zagouras, C. (2011). Prepare items for Large Scale Computer Based Assessment: Case study for Teachers' Certification on Basic Computer Skills. *Procedia-Social and Behavioral Sciences.* 29: 1189-1198.

[34] Clariana, R.B., Ross, S.M. & Morrison, G.R. (1991). The effects of different feedback strategies using computer-administered multiple choice questions as instruction. *Educational Technology Research and Development*. 39(2): 5-17.

[35] Clark, A. (2003). *Natural-Born Cyborgs: Why Minds and Technologies Are Made to Merge*. Oxford, UK : Oxford University Press.

[36] Cohen, L., Manion, L. & Morrison, K. (2011). *Research methods in Education.* 7th Edition. New York, USA: Routledge.

[37] Collins, A., Joseph, D. & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the Learning Sciences*. 13(1): 15–42.

[38] Conole, G. & Warburton, B. (2005). A review of computer-assisted assessment. *ALT-J, Research in Learning Technology*. 13(1):17-31.

[39] Cook, J. & Jenkins, V. (2010). Getting started with e-Assessment. *Project report. Bath: University of Bath*. Available from http://opus.bath.ac.uk/view/year/2010.html. [Accessed: 10th May 2014].

[40] Costagliola, G. & Fuccella, V. (2009). Online testing, current issues and future trends. *Journal of e-Learning and Knowledge Society (Je-LKS)*. 5(3): 79-90.

[41] CourseCompass assessment tool. Pearson Assessments. Available from www.coursecompass.com [Accessed: 8 September 2010].

[42] Cox, M.J. (2013). Formal to informal learning with IT: research challenges and issues for e-learning. *Journal of Computer Assisted Learning - Special Issue: Knowledge Transformation, Design and Technology*. 29(1): 85–105.

[43] Creswell, J.W. (2009). *Research Design: Qualitative, Quantitative and Mixed-Methods Approaches*. California, USA: SAGE Publications.

[44] Davies, P. (2001). *Computer Aided Assessment MUST be more than multiple-choice tests for it to be academically credible?* 5th CAA Conference. Loughborough: Loughborough University.

[45] de Villiers, M.R. (2012a). Models for Interpretive Information Systems Research, Part 1: IS Research, Action Research, Grounded Theory – A Meta-Study and Examples. In: *Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems*. M. Mora, O. Gelman, A. Steenkamp & M.S. Raisinghani (eds): 222-237. IGI Global, USA.

[46] de Villiers, M.R. (2012b). Models for Interpretive Information Systems Research, Part 2: Design Research, Development Research, Design-Science Research, and Design-Based Research – A Meta-Study and Examples. In: *Research Methodologies, Innovations and Philosophies in Software Systems Engineering and Information Systems*. M. Mora, O. Gelman, A. Steenkamp & M.S. Raisinghani (eds): 238-255. IGI Global, USA.

[47] de Villiers, M.R. & Harpur, P.A. (2013). Design-based research – the educational technology variant of design research: Illustrated by the design of an m-learning environment. *Proceedings of SAICSIT 2013, Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists*: 252-261. ACM International Conference Proceedings Series. South Africa: East London.

[48] Dempster, E.R. (2007). Textual strategies for answering multiple choice questions among South African students: what can we learn from TIMSS 2003? *African Journal of Research in SMT Education*. 11(1): 47-60.

[49] Derczeni, R. & Rogozea, l. (2011). Evaluation of Students Results After Computer Assessment. *Recent Researches in Educational Technology.* Available from http://www.wseas.us/e-library/conferences/2011/Iasi/EDUTE/EDUTE-26.pdf [Accessed: 1 April 2014].

[50] Deutsch, T., Herrmann, K., Frese, T. & Sandholzer, H. (2012). Implementing computer-based assessment – A web-based mock examination changes attitudes. *Computers & Education*. 58(4): 1068-1075.

[51] Douglas, K. M. & McGarty, C. (2001). Identifiability and self-presentation: Computer-mediated communication and intergroup interaction. *British Journal of Social Psychology.* 40(3): 399-416.

[52] Duarte, P., Nunes, I., Neto, J. P. & Chambel, T. (2006). Enhancing modularity and feedback in computer aided assessment. In: *Computing, 2006*. CIC'06, 15th International Conference: 240-246. IEEE.

[53] Dube, T. & Ma, M. (2011). Design of a flexible e-assessment system to accommodate diverse learning styles: a mathematics case study. Edulearn11 Proceedings: 6852-6860. *3rd International Conference on Education and New Learning Technologies*. Spain: Barcelona.

[54] Ebner, M. (2007). E-Learning 2.0=e-Learning 1.0 + Web 2.0. *Proceedings of the Second International Conference on Availability, Reliability and Security (ARES '07)*: 1235-1239.

[55] Elliott, J. (1991). Action Research for Educational Change. Buckingham: Open University Press.

[56] Elstein, A.S. (1993). Beyond multiple choice questions and essays: the need for a new way to assess clinical competence. *Academic Medicine.* 68 (4): 244.

[57] Engelbrecht, J. & Harding, A. (2003). E-assessment in mathematics: Multiple assessment formats. *New Zealand Journal of Mathematics*. 32: 57–66.

[58] ExamView Test Generator. Pearson Assessments. Available at http://www.formative-assessments.com/formative/examview/index.htm [Accessed: 8 September 2010].

[59] EzTest E-assessment tool. McGraw Hill. Available at www.eztestonline.com [Accessed: 8 September 2010].

[60] Falchikov, N. (2013). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education.* London: Routledge.

[61] Farthing, D.W., Jones, D.M. & McPhee, D. (1998.) Permutational multiple choice questions: an objective and efficient alternative to essay-type examination questions. *ACM SIGCSE Bulletin*. 30(3): 81-85.

[62] Fenderson, B.A., Damjanov, I., Robeson, M.R., Veloski, J.J. & Rubin, E. (1997). The virtues of extended matching and uncued tests as alternatives to multiple choice questions. *Human Pathology*. 28(5): 526-532.

[63] Feng, M., Heffernan, N. & Koedinger, K. (2009). Addressing the assessment challenge with an online system that tutors as it assesses. *User Modelling and User-Adapted Interaction*. 19(3): 243-266.

[64] Fielding, A. & Bingham, E. (2003). Tools for Computer Aided Assessment. *Learning and Teaching in Action.* 2(1).

[65] Fremont, D. J. & Jones, B. (1994). Computerized test question database project final report. Testing Software: a review. *New Currents 1.1*. Available at http://www.ucalgary.ca/pubs/Newsletters/Currents/Vol1.1/TestingSoftware.html. [Accessed: 8 September 2010].

[66] Gilles, J. L., Detroz, P. & Blais, J. G. (2012). An international online survey of the practices and perceptions of higher education professors with respect to the assessment of learning in the classroom. *Assessment & Evaluation in Higher Education.* 36(6): 719-733.

[67] Govender, D. (2003). *Cluster for Online Learning Environments (COLE).* Seminar Series No 1. Automated Assessment. Available from http://cole.unisa.ac.za/content/automated%20assessment.pdf [Accessed: 1 April 2014].

[68] Harper, R. (2003). Correcting computer-based assessments for guessing. *Journal of Computer Assisted Learning.* 19(1): 2-8.

[69] Harrington, C. F. & Reasons, S. G. (2005). Online student evaluation of teaching for distance education: A perfect match. *The Journal of Educators Online.* 2(1): 1-12.

[70] Hasibuan, Z.A. & Santoso, H.B. (2005). The use of e-learning towards a new learning paradigm: Case study student centred e-learning environment at Faculty of Computer Science-University of Indonesia. *Fifth IEEE International Conference on Advanced Learning Technologies*. IEEE Computer Society: 1026-1030.

[71] Hay, K.E, Kim, B. & Roy, T.C. (2005). Design-based research: more than formative assessment? An account of the virtual solar system project. *Educational Technology*. January-February 2005.

[72] Hepplestone, S. & Helm, P. (2003). Strategies for coping with unexpected uptake of CBA. In: Christie, J. (ed.) *Proceedings of 7th CAA Conference 2003: 159-167.* Loughborough: Loughborough University.

[73] Hevner, A. & Chatterjee, S. (2010). Design-science research in information systems. Springer US: 9-22.

[74] Hevner, A. R., March, S. T., Park, J. & Ram, S. (2004). Design science in information systems research. *MIS Quarterly.* 28(1): 75-105.

[75] Hodson, P., Saunders, D. & Stubbs, G. (2002). Computer-assisted assessment: Staff viewpoints on its introduction within a new university. *Innovations in Education and Teaching International*. 39(2): 145-152.

[76] Hofstee, E. (2006). *Constructing a good dissertation – a practical guide to finishing a Masters, MBA or PhD on schedule*. Sandton, South Africa: EPE Publishers.

[77] Honarmand, M. (2009). Computer Aided Assessment and its Situation in Higher Education: A Case Study from the Islamic Azad University in Iran. *2009 International Conference on Information Management and Engineering*. IEEE: 736.

[78] Horton, W. (2000). *Designing Web-based Training*. 1st Edition. USA. John Wiley & Sons Inc.

[79] Horton, W. & Horton, K. (2003). *E-Learning Tools and Technologies*. 1st edition. USA. Wiley.

[80] Hot Potatoes assessment tool. Half-Baked Software Inc. Available at www.hotpot.uvic.ca [Accessed: 8 November 2014].

[81] JISC. (2007). Effective Practice with e-Assessment: An overview of technologies, policies and practice in further and higher education. Available at

http://www.jisc.ac.uk/media/documents/themes/elearning/effpraceassess.pdf [Accessed: 14 October 2014].

[82] JISC. (2007). e-Assessment Glossary (extended). Available at http://www.jisc.ac.uk/uploaded_documents/eAssess-Glossary-Extended-v1-01.pdf [Accessed: 14 October 2014].

[83] Jordan, S. (2011). Using interactive computer-based assessment to support beginning distance students of science. *Open Learning.* 26(2): 147-164.

[84] Jordan, S. (2013). E-assessment: Past, present and future. *New Directions.* 9(1): 87-106.

[85] Kadhi, T. (2004). Online assessment: A study of the development and implementation of a formative online diagnostic tool in a college developmental mathematics course. *Paper presented at the annual meeting of the Southwest Educational Research Association, 5--7 February 2004*. Dallas: Texas A and M University.

[86] Karl, M., Graef, F., Eitner, S., Wichmann, M., Holst, S. & Beck, N. (2011). Student attitudes towards computer-aided testing. *European Journal of Dental Education*. 15(2): 69-72.

[87] Khan, B.H. (1997). *Web-Based Instruction*. New Jersey: Educational Technology Publications Ltd: 353-356.

[88] Khan, S., Davies, D.A. & Gupta, J.K. (2001). Formative self-assessment using multiple true-false questions on the Internet: feedback according to confidence about correct knowledge. *Medical Educator*. 23(2): 158-163.

[89] Khare, A. & Lam, H. (2008). Assessing Student Achievement and Progress with Online Examinations: Some Pedagogical and Technical Issues. *International Journal on E-Learning, 7*(3), 383-402. Chesapeake, VA: AACE.

[90] Khedo, K.K. (2005). Computer-assisted assessment system at the University of Mauritius. *IEEE 3rd International Conference on Computational Cybernetics*. ICCC 2005: 187.

[91] Kuechler, W.L. & Simkin, M.G. (2003). How well do multiple choice tests evaluate student understanding in computer programming classes? *Journal of Information Systems Education.* 14(4): 389-400.

[92] Laborda, G. & Royo, M. (2008). Does it Pay to Invest in Computer Based Testing Technology? Realities to Implement an Internet Based University Entrance Examination (iB PAU). *Paper*

presented at International CALL Research Conference "Practice-Based & Practice-Oriented CALL Research".

[93] Lambert, D. & Lines, D. (2013). *Understanding assessment: Purposes, perceptions, practice.* London: Routledge.

[94] Lambert, G. (2004). *What is Computer Aided Assessment and how can I use it in my teaching.* Learning and Teaching Unit Briefing Paper. Canterbury Christ Church University College. Available at http://www.canterbury.ac.uk/Support/learning-teaching-enhancement-unit/Resources/Documents/BriefingNotes/Blackboard.pdf [Accessed: 1 April 2014].

[95] Laumer, S. & von Stetten A. (2009). E-assessment. *Business & Information Systems Engineering*. 3: 263-265.

[96] Leshem, S. & Trafford, V. (2007). Overlooking the conceptual framework. *Innovations in education and Teaching International*. 44(1): 93-105.

[97] Lewis, D. J. & Sewell, R. D. (2007). Providing formative feedback from a summative computer-aided assessment. *American Journal of Pharmaceutical Education*. 71(2).

[98] Liu, L. (2014). Computer Information Technology Education Site Based on System Analysis. In: *Proceedings of the 9th International Symposium on Linear Drives for Industry Applications*. 1: 509-514. Berlin, Heidelberg: Springer.

[99] Lowry, R. (2005). Computer aided self-assessment – an effective tool. *Chemistry Education Research and Practice*. 6(4): 198-203.

[100] Luckett, K. & Sutherland, L. (2000). *Assessment practices that improve teaching and learning. Improving Teaching and Learning in Higher Education: A Handbook for Southern Africa*. Johannesburg, Witwatersrand. University Press & The Higher Education Research and Development Society of Australasia: 98-130.

[101] Macintosh, A. & Whyte, A. (2008). Towards an evaluation framework for eParticipation. *Transforming Government: People, Process and Policy*. 2(1): 16-30.

[102] Malau-Aduli, B. S., Assenheimer, D., Choi-Lundberg, D. & Zimitat, C. (2013). Using computer-based technology to improve feedback to staff and students on MCQ assessments. *Innovations in Education and Teaching International.* (ahead-of-print): 1-13.

[103] March, S.T. & Smith, G.F. (1995). Design and natural science research on information technology. *Decision Support Systems.* 15: 251-266.

[104] Maree, K. (2012). *First steps in research.* Pretoria, South Africa: Van Schaik Publishers.

[105] Maurice, S.A. & Day, R.L. (2004). Online Testing Technology: Important Lessons Learned. *International Journal of Engaging Education.* 20(1).

[106] McAlpine, M. (2002*). Principles of assessment*. E-assessment Centre, University of Luton.

[107] McCoubrie, P. (2004). Improving the fairness of multiple choice questions: a literature review. *Medical Educator.* 26(8): 709-712.

[108] Mendes, A. C., Curto, S. & Coheur, L. (2011). Bootstrapping Multiple-Choice Tests with The-Mentor. In: *computational linguistics and intelligent text processing*: 451-462. Heidelberg: Springer.

[109] Merriam, S.B. (2002). *Qualitative Research in Practice – examples for discussion and analysis*. San Francisco, USA: Jossey-Bass.

[110] Merriam, S.B. (2009). *Qualitative Research – a guide to design and implementation.* San Francisco, USA: Jossey-Bass.

[111] Messing, J. (2004). Online examinations in an academic environment: A case study at Charles Stuart University. In: *Proceedings of the Fifth International Conference on Information Technology Based Higher Education and Training*. New York, USA: IEEE.

[112] Miles, M.B. & Huberman, A.M. (1994). *An expanded sourcebook: Qualitative Data Analysis*. California, USA: SAGE Publications.

[113] Miller, P.A. (2012). Use of Computer Based Assessment Strategies. *Presented at the E-Learning Update*, August 2012, Johannesburg.

[114] Mitkov, R. & Ha, L.A. (2003). Computer-aided generation of multiple-choice tests. In: *Proceedings of the HLT-NAACL 03 workshop on Building educational applications using natural language processing*. 2(17). Stroudsbury PA: Association for Computational Linguistics.

[115] Mitkov, R., Ha, L.A. & Karamanis, N. (2006). A computer-aided environment for generating multiple-choice test items. *Natural Language Engineering*. 12(2): 177-194. New York: Cambridge University Press.

[116] Mogey, N. (2011). What is it that is really acting as a barrier to widespread use of summative e-assessment in UK higher education. *International Journal of e-Assessment.* 1(1).

[117] Moskal, P., Dziuban, C., Upchurch, R., Hartman, J. & Truman, B. (2006). Assessing online learning: What one university learned about student success, persistence, and satisfaction. *Peer Review*. 8(4): 26-29.

[118] Mostert, E., de Bruyn, E. & Pretorius, G. (2012). E-assessment – using computer based testing to enhance learning. *Paper presented at the E-Learning Update*. August 2012. Johannesburg.

[119] Mouton, J. (2008). *How to succeed in your Master's and Doctoral Studies – a South African guide and Resource book.* Pretoria, South Africa: Van Schaik Publishers.

[120] Myers, M.D. (1997). Qualitative Research in Information Systems. *Management Information Systems Quarterly*, *21*, 241-242.

[121] Nelson, G.E. (1998). On-line Evaluation: Multiple Choice, Discussion Questions, Essay, and Authentic Projects. Available from http://leahi.kcc.hawaii.edu/org/tcon98/paper/nelson.html [Accessed: 4 January 2014].

[122] Nicol, D. (2007). E-assessment by design: using multiple-choice tests to good effect. *Journal of Further and Higher Education.* 31(1): 53-64.

[123] Oates, B.J. (2010). *Researching Information Systems and Computing.* 5th Edition. London, UK: Sage Publications.

[124] Offerman, P., Blom, S., Schonherr, M. & Bub, U. (2010). Artifact types in Information Systems Design Science – a Literature Review. In: Winter, R., Zhao, J.L. & Aier, S. (eds). *DESRIST 2010*. 337-355. Heidelberg: Springer-Verlag.

[125] Olivier, M. (2004). *Information Technology Research*. 2nd Edition. Pretoria, South Africa: Van Schaik Publishers.

[126] Osika, E.R. (2006). Assessing Student Learning Online: It's More Than Multiple Choice. *Online Cl@ssroom:Ideas for Effective Online Instruction.* 29: 208.

[127] Parshall, C. G., Spray, J. A., Kalohn, J. C. & Davey, T. (2002). Issues in innovative item types. *Practical considerations in computer-based testing*: 70-91. New York: Springer.

[128] Passey, D. (2011). Managing quality education–identifying the learning needs of the individual, then satisfying them. *Information Technology and Managing Quality Education* : 1-10. Heidelberg: Springer.

[129] Patel, J. B., Kothari, R. G. & Makwana, S. V. (2013). E assessment - An Easy Way to Evaluate Student. In: *National Conference on Assessment Practices in Schools*. Available from:

http://www.riemysore.ac.in/conference/index.php/Riemaa/ncaps/paper/view/103 [Accessed: 1 April 2014].

[130] Pattinson, S. (2004). The use of CAA for formative and summative assessment – student views and outcomes. In: *Proceedings of the 8th CAA Conference.* Loughborough: Loughborough University.

[131] Peffers, K., Tuunanen, T., Rothenberger, M.A. & Chatterjee, S. (2008). A design-science research methodology for Information Systems Research. *Journal of Management Information Systems.* 24(3): 45-77.

[132] Pittenger, A. L. & Lounsbery, J. L. (2011). Student-generated questions to assess learning in an online orientation to pharmacy course. *American Journal of Pharmaceutical Education.* 75(5).

[133] Prestera, G.E., Clariana, R. & Peck, A. (2005). Memory-context effects of screen color in multiple-choice and fill-in tests. *Journal of Educational Multimedia and Hypermedia*. 14(4): 415.

[134] Pretorius, G.J., Mostert, E. & de Bruyn, E. (2007). Local innovation: development of a computer-based testing system. In: *9th Annual Conference on World Wide Web Applications*. Johannesburg. South Africa.

[135] Roberts, T.S. (2006). The use of multiple choice tests for formative and summative assessment. In: *Proceedings of the 8th Australian conference on Computing Education*. Australian Computer Society, Inc. 5: 175.

[136] Roediger, H. L., Agarwal, P. K., McDaniel, M. A. & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology.* 17: 382–395.

[137] Rovai, A.P. (2000). Online and traditional assessments: what is the difference? *The Internet and Higher Education.* 3:141-151.

[138] Rusman, E., Boon, J., Martínez-Monés, A., Rodríguez-Triana, M. J. & Retalis, S. (2013). *Towards the Use of New Methods for Formative e-assessment of 21st Century Skills in Schools*. Available from http://dspace.ou.nl/bitstream/1820/5022/1/TEFA_manuscript.pdf [Accessed: 1 April 2014].

[139] Russell, C. (2009). A systemic framework for managing e-learning adoption in campus universities: individual strategies in context. *Association for Learning Technology Journal.* 17(1): 3-19.

[140] SAM assessment tool. Cengage Course Technology. Available from samcentral.course.com [Accessed: 8 September 2010].

[141] Sangi, N.A. (2008). Electronic assessment issues and practices in Pakistan: a case study. *Learning, Media and Technology.* 33(3): 191-206.

[142] Scalise, K. & Gifford, B. (2006). Computer-Based Assessment in E-Learning: A Framework for Constructing - Intermediate Constraint Questions and Tasks for Technology Platforms. *Journal of Technology, Learning, and Assessment*. 4: 45.

[143] Scalter, N. & Howie, K. (2003). User requirements of the ultimate online assessment engine. *Computers & Education*. 40(3): 285-306.

[144] Schuwirth, L., Vleuten, C.P.M. & Donkers, H. (1996). A closer look at cueing effects in multiple choice questions. *Medical Education*. 30(1): 44-49.

[145] Seale, J. (2002). Using CAA to support student learning. *LTSN Generic Centre Starter Guide*. 4.

[146] Shulman, T.D. (2005). Perception questionmark. *Internet and Higher Education.* 8: 263-267.

[147] Sim, G., Holifield, P. & Brown, M. (2004). Implementation of computer assisted assessment: lessons from the literature. *ALT-J, Research in Learning Technology*. 12(3).

[148] Simkin, M.G. & Kuechler, W.L. (2005). Multiple-Choice Tests and Student Understanding: What Is the Connection? *Decision Sciences Journal of Innovative Education*. 3(1): 73-98.

[149] Simon, H.A. (1981). *The sciences of the artificial*. 2^{nd} Edition. Cambridge, MA: MIT Press.

[150] Singh, U.G. & de Villiers, M.R. (2010). Establishing the current extent and nature of usage of Online Assessment Tools in Computing-related Departments at South African Tertiary Institutions. In: *SACLA 2010 Conference Proceedings*. Pretoria, South Africa.

[151] Singh, U.G. & De Villiers, M.R. (2012). Investigating the use of different kinds of multiple-choice questions in electronic assessment. Progressio 34(3): 125-143.

[152] Souali, K., Afia, A. E., Faizi, R. & Chiheb, R. (2011). A new recommender system for e-learning environments. In: *Multimedia Computing and Systems (ICMCS), 2011*. IEEE International Conference: 1-4.

[153] Swaffield, S. (2008). *Unlocking Assessment: Understanding for Reflection and Application.* London: Routledge.

[154] Testa, A.M. (2008). Assessment of Student Learning through an Online, Competency-Based University. *Assessment Update.* 3.

[155] The Design-Based Research Collective. (2003). Design-based research: an emerging paradigm for educational enquiry. *Educational Researcher*. 32(1) January/February 2003.

[156] Tsai, S. & Machado, P. (2002). E-learning, Online Learning, Web-based Learning, or Distance Learning: Unveiling the Ambiguity in Current Terminology. *e-Learn Magazine*. Available from http://www.elearnmag.org/subpage.cfm?section=best_practice [Accessed: 4 January 2014].

[157] Tsintsifas, A. (2002). A framework for the computer-based assessment of diagram-based coursework. *Computer Science and Information Technology*. Loughborough: Loughborough University.

[158] Turban, E., King, D., Lee, J., Liang, T.P. & Turban, D. (2010). *Electronic Commerce 2010*. 6th Edition. New Jersey. Pearson Education Inc.

[159] Valenti, S., Cucchiarelli, A. & Panti, M. (2002). Computer based assessment systems evaluation via the ISO90126 quality model. *Journal of Information Technology Education.* 1(3): 157–175.

[160] Venkatesh, V., Brown, S. A. & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly.* 37(1): 21-54.

[161] Ventouras, E., Triantis, D., Tsiakas, P. & Stergiopoulos, C. (2010). Comparison of examination methods based on multiple choice questions and constructed-response questions using personal computers. *Computers & Education.* 54(2): 455-461.

[162] Walker, D.J., Topping, K. & Rodrigues, S. (2008). Student reflections on formative e-assessment: expectations and perceptions. *Learning, Media and Technology*. 33(3): 221-234.

[163] Wang, F. & Hannafin, M. (2005). Design-Based Research and Technology-Enhanced Learning Environments. *Educational Technology Research and Development.* 53(4).

[164] Ware, J., Kattan, T. E., Siddiqui, I., & Mohammed, A. M. (2014). The perfect MCQ exam. Journal of Health Specialties, 2(3): 94.

[165] Waring, A., Farthing, C.B. & Kidder-Ashley, P. (1999). Impulsive Response Style Affects Computer-Administered Multiple-Choice Test Performance. *Journal of Instructional Psychology.* 26(2): 121-128.

[166] Weerakoon, G.L.P. (2001). The role of e-assessment in health professional education: a comparison of student performance in computer-based and paper-and-pen multiple-choice tests. *Medical Educator*. Available from www.informahealthcare.com [Accessed: 4 January 2014].

[167] Wen, M.L. & Tsai, C.C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education*. 51(1): 27-44.

[168] White, R.L., Myers, S. & Earl Sr, A.W. (2008). A Description of the Computer Assisted Assessment Program in University Elementary Algebra at Norfolk State University. Available from http://files.eric.ed.gov/fulltext/ED503057.pdf [Accessed 1 April 2014].

[169] Whitelock, D., Gilbert, L. & Gale, V. (2011). *Technology Enhanced Assessment and Feedback: How is evidence-based literature informing practice?* Available from http://oro.open.ac.uk/29081/ [Accessed 1 April 2014].

[170] Wood, E.J. (2003). What are extended matching sets questions. *Bioscience Education eJournal*. 1(1): 1-2.

[171] Wu, Y. & Di, C. (2009). Application Oriented Teaching Method to the Course of College Computer Fundamentals. In: *Proceedings of the 2009 First International Workshop on Education Technology and Computer Science.* 2: 334-336. IEEE Computer Society.

[172] Yang, J.Y. (2008). Teaching Academic Courses Online: An Assessment of San Diego Miramar College Students. *Online Submission:* 9. Available from http://files.eric.ed.gov/fulltext/ED500925.pdf [Accessed 1 April 2014].

[173] Yonker, J. E. (2011). The relationship of deep and surface study approaches on factual and applied test-bank multiple-choice question performance. *Assessment & Evaluation in Higher Education*. 36(6): 673–686.

[174] Yu, Y. & Peng, X. (2009). Discussion of Computer Basis Education Reform and Innovation in Colleges and Universities. In: *Proceedings of the 2009 First International Workshop on Education Technology and Computer Science.* 3: 367. IEEE Computer Society.

[175] Yusof, M.M., Papazafeiropoulou, A., Paul, R.J. & Stergioulas, L.K. (2008). Investigating evaluation frameworks for health information systems. *International Journal of Medical Informatics*. 77: 377-385.

[176] Zuber-Skerritt, O. (1992). *Action research in higher education: Examples and reflections*. London: Kogan Page.

APPENDICES

Please refer to the CD attached to the back cover of this thesis, for the appendices listed below.

APPENDIX A: UNISA ethical clearance documentation

A1: 2009 Ethical clearance for Study 1
A2: 2011 Addendum for Studies 2 and 3
A3: 2013 Addendum for Studies 4 to 6

APPENDIX B: UKZN ethical clearance for Study 1

APPENDIX C: Questionnaire 1 for Study 1

APPENDIX D: Documents for Interviews in Study 2

APPENDIX D1: Personal interviews schedule for Study 2
APPENDIX D2: Informed consent for personal interviews in Study 2 (anonymized example)
APPENDIX D3: Telephonic interviews schedule for Study 2

APPENDIX E: Questionnaire 2 for Study 3

APPENDIX F: Documents for Study 4a - Pilot Study

APPENDIX F1: Instrument for Study 4a - Pilot Study (MS Word Version)
APPENDIX F2: Instrument for Study 4a - Pilot Study (Survey Monkey version)
APPENDIX F3: Pilot Study questionnaire

APPENDIX G: Instrument for Study 4b - Evaluation Study

APPENDIX H: Documents for Study 4c - Proof of Concept Study

APPENDIX H1: Instrument for Study 4c - Proof of Concept Study
APPENDIX H2: Proof of Concept Study questionnaire

APPENDIX I: Instrument for Study 4d - Application Study

APPENDIX J: Documents for Studies 5 and 6: SEAT and e-SEAT

APPENDIX J1: e-SEAT instrument
APPENDIX J2: SEAT questionnaire
APPENDIX J3: e-SEAT questionnaire

APPENDIX K: e-SEAT instruction file

APPENDIX L: Publications from this research

APPENDIX L1: SACLA Conference Paper
APPENDIX L2: Progressio Journal Article