ALTERNATIVE MEASUREMENT INSTRUMENTS FOR THE EVALUATION OF LEARNER ACHIEVEMENT IN BUSINESS MANAGEMENT AT THE TECHNIKON SA

by

JOHAN FRANS VAN KOLLER

submitted in accordance with the requirements for the degree of

DOCTOR OF EDUCATION

in the subject

DIDACTICS

at the

UNIVERSITY OF SOUTH AFRICA

PROMOTER: PROF. L J VAN NIEKERK

NOVEMBER 1997
"I tell you the truth, if you have faith as small as a mustard seed, you can say to this mountain, 'Move from here to there' and it will move. Nothing will be impossible for you.”

(Matthew 17: 20)
ACKNOWLEDGEMENTS

My sincere thanks to the following people:

- Our heavenly Father who made it possible for me to start and complete this research.

- My promoter, Professor LJ van Niekerk, for his challenging and inspiring guidance and supervision.

- Mrs Hester Mey for her assistance in the proof-reading and editing of this document.

- Miss Karen Hendricks for the transcription of the focus group interviews.

- My wife, Naomi, for all the typing she has done and for her loyal support.

- My daughters, Melisha, Racha and Nikita, whom I have neglected very much.

- Everyone who motivated and encouraged me to continue my research.
SUMMARY

This research is mainly concerned with determining whether the existing evaluation methods/measurement instruments used in Business Management at Technikon SA, are still suitable, adequate and acceptable enough for continued use, or whether there is a need for the introduction of alternative methods.

The first objective in this research was to determine what the exact role and function of evaluation in the education process is. On the basis of literature, it was determined that evaluation, and by implication measurement, is a crucial component of the education process and that no real teaching or learning can occur without it.

Second, an attempt was made to determine what the nature and extent of the shortcomings of the existing measurement instruments were. Also on the basis of literature it was found that there are numerous and serious disadvantages in the existing measurement instruments.

Third, the researcher also attempted to determine whether any other (alternative) measurement instruments, in addition to the existing ones, were available. It was discovered that there are a variety of alternative evaluation/measurement instruments available.

Fourth, the measurement instruments used in Business Management were analysed and compared to the Business Management curriculum objectives to determine whether they were indeed suitable for the subject. It was found that, although a positive correlation seems to exist between the measurement instruments and the curriculum objectives, there were certain definite shortcomings in the existing measurement instruments in terms of the realization of some of the curriculum objectives.

Fifth, focus group interviews were used as a research method to get an understanding of the views of TSA learners on the aspect of evaluation and the methods/instruments currently used. Most of the participants expressed a need for more practical types of
measurement/evaluation. However, many participants felt the written examination should not be abolished entirely, but be used in conjunction with other more practical methods.

The research concluded with a number of recommendations based on the findings in this research.

Key terms:

Evaluation
Summative evaluation
Measurement
Measurement instruments
Alternative measurement instruments
Performance-based measurement
Learners
Achievement
Qualitative research
Focus group interviews
TABLE OF CONTENTS

CHAPTER 1

PROBLEM STATEMENT AND AIMS OF RESEARCH

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>BECOMING AWARE OF THE PROBLEM</td>
<td>2</td>
</tr>
<tr>
<td>1.3</td>
<td>STATING THE PROBLEM</td>
<td>4</td>
</tr>
<tr>
<td>1.4</td>
<td>AIMS AND PURPOSES OF THIS RESEARCH</td>
<td>7</td>
</tr>
<tr>
<td>1.5</td>
<td>RESEARCH METHODOLOGY</td>
<td>8</td>
</tr>
<tr>
<td>1.5.1</td>
<td>Literature review</td>
<td>8</td>
</tr>
<tr>
<td>1.5.2</td>
<td>The qualitative research method</td>
<td>10</td>
</tr>
<tr>
<td>1.5.2.1</td>
<td>What is qualitative research?</td>
<td>10</td>
</tr>
<tr>
<td>1.5.2.2</td>
<td>Why qualitative research?</td>
<td>11</td>
</tr>
<tr>
<td>1.5.2.3</td>
<td>How does qualitative research operate?</td>
<td>13</td>
</tr>
<tr>
<td>1.6</td>
<td>DEMARCATIONS AND LIMITATIONS</td>
<td>16</td>
</tr>
<tr>
<td>1.7</td>
<td>WHY BUSINESS MANAGEMENT?</td>
<td>17</td>
</tr>
<tr>
<td>1.8</td>
<td>CONCEPT DEFINITIONS</td>
<td>19</td>
</tr>
<tr>
<td>1.8.1</td>
<td>Evaluation</td>
<td>19</td>
</tr>
<tr>
<td>1.8.2</td>
<td>Summative evaluation</td>
<td>23</td>
</tr>
<tr>
<td>1.8.3</td>
<td>Measurement</td>
<td>23</td>
</tr>
<tr>
<td>1.8.4</td>
<td>Measurement instruments</td>
<td>25</td>
</tr>
<tr>
<td>1.8.5</td>
<td>Alternative measurement instruments</td>
<td>27</td>
</tr>
<tr>
<td>1.8.6</td>
<td>Learner</td>
<td>28</td>
</tr>
<tr>
<td>1.8.7</td>
<td>Achievement</td>
<td>29</td>
</tr>
<tr>
<td>1.9</td>
<td>STRUCTURE OF THE RESEARCH PROGRAMME</td>
<td>30</td>
</tr>
<tr>
<td>1.10</td>
<td>SUMMARY</td>
<td>32</td>
</tr>
</tbody>
</table>
## A BROAD OVERVIEW OF THE PROCESSES OF EVALUATION AND MEASUREMENT

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>INTRODUCTION</td>
<td>33</td>
</tr>
<tr>
<td>2.2</td>
<td>EDUCATIONAL EVALUATION</td>
<td>34</td>
</tr>
<tr>
<td>2.2.1</td>
<td>The origin and historical development of educational evaluation</td>
<td>34</td>
</tr>
<tr>
<td>2.2.1.1</td>
<td>General development until 1950</td>
<td>34</td>
</tr>
<tr>
<td>2.2.1.2</td>
<td>Important role players in the history of educational evaluation</td>
<td>36</td>
</tr>
<tr>
<td>2.2.2</td>
<td>New trends in educational evaluation</td>
<td>41</td>
</tr>
<tr>
<td>2.2.3</td>
<td>The essential role of evaluation in the educational process</td>
<td>44</td>
</tr>
<tr>
<td>2.2.3.1</td>
<td>Evaluation as a prerequisite for the teaching process to succeed</td>
<td>45</td>
</tr>
<tr>
<td>2.2.3.2</td>
<td>Evaluation of the extent to which instructional objectives have been met</td>
<td>47</td>
</tr>
<tr>
<td>2.2.4</td>
<td>Evaluation and the curriculum</td>
<td>48</td>
</tr>
<tr>
<td>2.2.5</td>
<td>Various other purposes of educational evaluation</td>
<td>51</td>
</tr>
<tr>
<td>2.2.6</td>
<td>Educational evaluation models</td>
<td>53</td>
</tr>
<tr>
<td>2.2.7</td>
<td>The main forms of educational evaluation</td>
<td>61</td>
</tr>
<tr>
<td>2.3</td>
<td>EDUCATIONAL MEASUREMENT</td>
<td>62</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Measurement: The essential component of evaluation</td>
<td>62</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Basic forms of educational measurement</td>
<td>63</td>
</tr>
<tr>
<td>2.3.2.1</td>
<td>Norm-referenced testing</td>
<td>64</td>
</tr>
<tr>
<td>2.3.2.2</td>
<td>Criterion-referenced testing</td>
<td>64</td>
</tr>
<tr>
<td>2.3.2.3</td>
<td>Other test types</td>
<td>65</td>
</tr>
<tr>
<td>2.3.3</td>
<td>The essential requirements for a measurement instrument</td>
<td>67</td>
</tr>
<tr>
<td>2.3.3.1</td>
<td>Validity</td>
<td>68</td>
</tr>
<tr>
<td>2.3.3.2</td>
<td>Reliability</td>
<td>69</td>
</tr>
<tr>
<td>2.3.3.3</td>
<td>Accountability/Legitimacy</td>
<td>71</td>
</tr>
</tbody>
</table>
2.3.4 The exact nature of educational measurement
2.3.4.1 The process of measurement
2.3.4.2 What is measured in educational measurement?
2.3.4.3 How are attributes defined in educational measurement?
2.3.4.4 Difficulties in educational measurement as compared to physical measurement

2.4 IMPLICATIONS OF RECENT DEVELOPMENTS IN THE SOUTH AFRICAN EDUCATIONAL CONTEXT
2.4.1 The NQF and SAQA
2.4.1.1 What is the NQF and SAQA
2.4.1.2 The influence of the NQF on evaluation practices
2.4.1.3 The Education, Training and Development Practices Project
2.4.1.4 SERTEC
2.4.2 Outcomes-based Education (OBE)
2.4.2.1 What is OBE?
2.4.2.2 Curriculum 2005
2.4.3 Technology-Enhanced Learning Initiative

2.5 IMPLICATIONS OF RECENT SCIENTIFIC DEVELOPMENTS ON THE EVALUATION AND MEASUREMENT PROCESSES
2.5.1 What is post-modernism?
2.5.2 What are the implications of post-modernism on education?
2.5.3 What are the implications of post-modernism on the evaluation process?

2.6 SUMMARY
CHAPTER 3

CONVENTIONAL SUMMATIVE MEASUREMENT INSTRUMENTS

3.1 INTRODUCTION 102

3.2 WHAT IS MEANT BY CONVENTIONAL SUMMATIVE MEASUREMENT INSTRUMENTS? 103

3.3 A BASIC DESCRIPTION OF THE MOST COMMON CONVENTIONAL MEASUREMENT INSTRUMENTS 103

3.3.1 Different methods/techniques of measurement 103

3.3.2 The Standard Conventional Examination 109

3.3.2.1 Objective questions 110

3.3.2.2 Subjective questions 113

3.3.3 Essay-type Examinations 115

3.3.4 Structured Essay-type Questions 117

3.3.5 Written Case Studies 118

3.3.6 The Open-book Examination 119

3.3.7 The Pre-set Question (Seen Examination) 119

3.4 CONVENTIONAL MEASUREMENT INSTRUMENTS - A GLOBAL VIEW. 120

3.4.1 Older forms 122

3.4.1.1 The United States of America 122

3.4.1.2 England and Wales 127

3.4.1.3 France 128

3.4.1.4 Federal Republic of Germany 129

3.4.1.5 Sweden 129

3.4.1.6 Union of Soviet Socialist Republics 130

3.4.1.7 The People's Republic of China 131

3.4.1.8 Japan 131

3.4.1.9 Australia 133

3.4.1.10 Africa 133
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5.4.4 Reliability of the essay-type question</td>
<td>172</td>
</tr>
<tr>
<td>3.5.4.5 Suitability of the essay-type question</td>
<td>173</td>
</tr>
<tr>
<td>3.5.5 Scrutiny of the structured essay-type question</td>
<td>174</td>
</tr>
<tr>
<td>3.5.5.1 Advantages of the structured essay-type question</td>
<td>174</td>
</tr>
<tr>
<td>3.5.5.2 Disadvantages of the structured essay-type question</td>
<td>176</td>
</tr>
<tr>
<td>3.5.5.3 Validity of the structured essay-type question</td>
<td>177</td>
</tr>
<tr>
<td>3.5.5.4 Reliability of the structured essay-type question</td>
<td>177</td>
</tr>
<tr>
<td>3.5.5.5 Suitability of the structured essay-type question</td>
<td>177</td>
</tr>
<tr>
<td>3.5.6 Scrutiny of the written case study</td>
<td>178</td>
</tr>
<tr>
<td>3.5.6.1 Advantages of the written case study</td>
<td>178</td>
</tr>
<tr>
<td>3.5.6.2 Disadvantages of the written case study</td>
<td>178</td>
</tr>
<tr>
<td>3.5.6.3 Validity of the written case study</td>
<td>179</td>
</tr>
<tr>
<td>3.5.6.4 Reliability of the written case study</td>
<td>179</td>
</tr>
<tr>
<td>3.5.6.5 Suitability of the written case study</td>
<td>179</td>
</tr>
<tr>
<td>3.5.7 Scrutiny of the open-book examination</td>
<td>180</td>
</tr>
<tr>
<td>3.5.7.1 Advantages of the open-book examination</td>
<td>180</td>
</tr>
<tr>
<td>3.5.7.2 Disadvantages of the open-book examination</td>
<td>182</td>
</tr>
<tr>
<td>3.5.7.3 Validity of the open-book examination</td>
<td>183</td>
</tr>
<tr>
<td>3.5.7.4 Reliability of the open-book examination</td>
<td>183</td>
</tr>
<tr>
<td>3.5.7.5 Suitability of the open-book examination</td>
<td>184</td>
</tr>
<tr>
<td>3.5.8 Scrutiny of the pre-set question</td>
<td>185</td>
</tr>
<tr>
<td>3.5.8.1 Advantages of the pre-set question</td>
<td>185</td>
</tr>
<tr>
<td>3.5.8.2 Disadvantages of the pre-set question</td>
<td>186</td>
</tr>
<tr>
<td>3.5.8.3 Validity of the pre-set question</td>
<td>187</td>
</tr>
<tr>
<td>3.5.8.4 Reliability of the pre-set question</td>
<td>188</td>
</tr>
<tr>
<td>3.5.8.5 Suitability of the pre-set question</td>
<td>188</td>
</tr>
<tr>
<td>3.5.9 CONCLUSION</td>
<td>189</td>
</tr>
<tr>
<td>3.6 SUMMARY</td>
<td>190</td>
</tr>
</tbody>
</table>
CHAPTER 4

ALTERNATIVE EDUCATIONAL MEASUREMENT INSTRUMENTS

4.1 INTRODUCTION 191

4.2 WHAT IS MEANT BY ALTERNATIVE FORMS OF EVALUATION/MEASUREMENT? 191

4.2.1 Alternative measurement 192

4.2.2 Authentic measurement 192

4.2.3 Curriculum-based/Curriculum-embedded (performance) measurement 193

4.2.4 Dynamic measurement 193

4.2.5 Portfolios 194

4.2.6 Performance-based measurement 194

4.2.7 Conclusion 194

4.3 WHY THE NEED FOR ALTERNATIVE MEASUREMENT? 195

4.4 PERFORMANCE-BASED MEASUREMENT 197

4.4.1 Definition 197

4.4.2 Different types of performance-based measurement 198

4.4.3 Characteristics of performance-based measurement 200

4.4.4 Validity of performance-based measurement 203

4.4.5 Reliability of performance-based measurement 207

4.4.6 Accountability of performance-based measurement 208

4.4.7 General advantages of performance-based measurement 212

4.4.8 General disadvantages of performance-based measurement 213

4.5 A BRIEF SCRUTINY OF PERFORMANCE-BASED MEASUREMENT INSTRUMENTS 214

4.5.1 Actual performance measurement 214

4.5.1.1 Work-sample tests 215

4.5.1.2 Identification tests 219

4.5.1.3 Supervisor ratings 222

4.5.1.4 Peer ratings 224

4.5.1.5 Self-evaluation 227
4.5.2 Simulations

4.5.2.1 Simulated performance tests

4.5.2.2 Simulated identification tests

4.5.2.3 Written simulations

4.5.2.4 Management exercises

4.5.2.4.1 Role-playing

4.5.2.4.2 Simulated interviews

4.5.2.4.3 Fact-finding exercises

4.5.2.4.4 Case study

4.5.2.4.5 Leaderless group discussions

4.5.2.4.6 Other types

4.5.3 Observational measurement

4.5.3.1 General description

4.5.3.2 General advantages

4.5.3.3 General disadvantages

4.5.3.4 Different observational measurement instruments

4.5.3.4.1 Checklists

4.5.3.4.2 Rating scales

4.5.3.4.3 Anecdotal records

4.5.4 Oral measurement

4.5.4.1 Oral examinations

4.5.4.2 Interviews

4.5.4.3 Prepared presentations

4.5.5 Programme requirements

4.5.5.1 Personal records

4.5.5.2 Performance records

4.5.5.3 Apprenticeship/Internship

4.5.5.4 Questionnaires

4.6 A BRIEF SCRUTINY OF COMPUTER-BASED EVALUATION

4.6.1 Description of computer-based measurement

4.6.2 Advantages of computer-based measurement

4.6.3 Disadvantages of computer-based measurement
<table>
<thead>
<tr>
<th>4.6.4</th>
<th>Suitability for TSA conditions</th>
<th>293</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6.5</td>
<td>Suitability for Business Management</td>
<td>294</td>
</tr>
<tr>
<td>4.6.6</td>
<td>Financial implications</td>
<td>294</td>
</tr>
<tr>
<td>4.7</td>
<td>CONCLUSION</td>
<td>295</td>
</tr>
<tr>
<td>4.8</td>
<td>SUMMARY</td>
<td>297</td>
</tr>
</tbody>
</table>
CHAPTER 5

BUSINESS MANAGEMENT CURRICULUM VERSUS MEASUREMENT INSTRUMENTS

5.1 INTRODUCTION 298

5.2 RESEARCH METHODOLOGY 298

5.2.1 Survey and questionnaire 299

5.2.2 Document review 300

5.3 MEASUREMENT INSTRUMENTS IN BM AT SA TERTIARY EDUCATION INSTITUTIONS 301

5.4 MEASUREMENT INSTRUMENTS IN BM AT TSA 309

5.5 SERTEC PRINCIPLES CONCERNING THE CURRICULUM 310

5.6 THE BM CURRICULUM VERSUS MEASUREMENT INSTRUMENTS 312

5.6.1 Findings from the BM I curriculum and measurement instruments 313

5.6.1.1 Findings from an analysis of the BM I curriculum 313

5.6.1.2 Findings from an analysis of BM I measurement instruments 314

5.6.1.3 Findings from a comparison between BM I curriculum objectives and measurement instruments 317

5.6.1.4 Findings from a comparison of BM I curriculum objectives and measurement instruments with SERTEC prescriptions 317

5.6.2 Findings from the BM II curriculum and measurement instruments 318

5.6.2.1 Findings from an analysis of the BM II curriculum 318

5.6.2.2 Findings from an analysis of BM II measurement instruments 320

5.6.2.3 Findings from a comparison between BM II curriculum objectives and measurement instruments 323

5.6.2.4 Findings from a comparison of BM II curriculum objectives and measurement instruments with SERTEC prescriptions 323

5.6.3 Findings from the BM III curriculum and measurement instruments 323

5.6.3.1 Findings from an analysis of the BM III curriculum 323
5.6.3.2 Findings from an analysis of BM III measurement instruments 327
5.6.3.3 Findings from a comparison between BM III curriculum objectives and measurement instruments 331
5.6.3.4 Findings from a comparison of BM III curriculum objectives and measurement instruments with SERTEC prescriptions 332
5.7 SUMMARY OF FINDINGS 333
5.8 SUMMARY 333
CHAPTER 6

VIEWS AND PERCEPTIONS OF LEARNERS IN BM AT TSA

6.1 INTRODUCTION 335
6.2 RESEARCH METHODOLOGY 335
6.2.1 Introduction 335
6.2.2 Interviews 336
6.2.2.1 What is an interview? 336
6.2.2.2 The purpose of interviews 336
6.2.2.3 Advantages of interviews 337
6.2.2.4 Disadvantages of interviews 337
6.2.2.5 Different types of interviews 338
6.2.3 Focus group interviews 339
6.2.3.1 What are focus groups? 339
6.2.3.2 The purpose of focus groups 339
6.2.3.3 Characteristics of focus group interviews 340
6.2.3.4 Advantages of focus groups 340
6.2.3.5 Disadvantages of focus groups 342
6.2.4 Seminar 343
6.3 STRUCTURE OF RESEARCH 345
6.3.1 Introduction 345
6.3.2 The size and number of focus groups 347
6.3.3 Selection of participants 349
6.3.4 Types of questions 351
6.3.5 Tape recordings of interviews 351
6.3.6 Duration of sessions 352
6.4 ANALYSIS OF DATA FROM FOCUS GROUPS 352
6.5 FINDINGS BASED ON THE ANALYSIS OF THE TRANSSCRIPTS 353
6.5.1 Main categories 353
6.5.1.1 Category 1: The importance of evaluation 354
6.5.1.2 Category 2: The purpose of evaluation 355
6.5.1.3 Category 3: Difficulties/problems regarding the written examination 357
6.5.1.4 Category 4: Written examinations versus practical evaluation 360
6.5.2 Summary of findings from focus group interviews 365
6.6 FINDINGS OF THE ANALYSIS OF DATA FROM SEMINAR 367
6.6.1 Brief summary of papers 367
6.6.1.1 Van Niekerk 367
6.6.1.2 Frazer 368
6.6.1.3 Du Plessis 369
6.6.1.4 Gericke 369
6.6.1.5 Dresselhaus 370
6.6.2 Selection of main categories 370
6.6.2.1 Category 1: The purpose of evaluation in education 372
6.6.2.2 Category 2: Evaluation methods 374
6.6.2.3 Category 3: Future implications/requirements 376
6.6.3 Summary of findings from analysis of seminar papers 381
6.7 FINDINGS FROM A COMPARISON BETWEEN CATEGORIES OF INTERVIEWS AND SEMINAR 383
6.8 SUMMARY 384
CHAPTER 7

SUMMARY, MAIN FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

7.1 SUMMARY OF RESEARCH 386
7.2 A REFLECTION ON QUESTIONS POSED IN CHAPTER 1 388
7.3 SUMMARY OF MAIN FINDINGS 389
7.3.1 General findings 389
7.3.2 Specific findings from the document review in chapter 5 391
7.3.3 Specific findings from focus group interviews and seminar in chapter 6 392
7.3.3.1 Findings from an analysis of the transcripts of the focus group interviews 392
7.3.3.2 Findings from an analysis of the seminar papers 393
7.3.3.3 Findings from a comparison between categories of interviews and seminar 396
7.4 CONCLUSIONS AND RECOMMENDATIONS 396
7.4.1 Conclusion 1 397
7.4.2 Conclusion 2 397
7.4.3 Conclusion 3 398
7.4.4 Conclusion 4 399
7.4.5 Conclusion 5 400
7.4.6 Conclusion 6 401
7.4.7 Conclusion 7 402
7.5 FURTHER AREAS OF RESEARCH 402
7.6 CLOSING REMARKS 403
8. BIBLIOGRAPHY 405
9. ANNEXURES 427
ABBREVIATIONS

The following abbreviations were commonly used in this research:

TSA: TECHNIKON SOUTHERN AFRICA
BM: BUSINESS MANAGEMENT
CHAPTER 1
PROBLEM STATEMENT AND AIMS OF RESEARCH

1.1 INTRODUCTION

The phenomenon of evaluation appears to be a central “ingredient” in the total teaching and instruction process world-wide. Evaluation is regarded as one of the essential elements of education as a whole and, together with the setting of objectives (learning/curriculum) and the instruction process itself, it forms the backbone of teaching (Hopkins & Antes, 1990:39; Mehrens & Lehmann, 1991:29, 47; Cangelosi, 1990:4).

In a South Africa in midst of a process of transformation at political, social, cultural and educational level, almost everything is bound to be affected. The changes in the educational sphere are particularly radical as approximately 38 educational departments merged into a single department. These departments had previously racially separated schools which are currently open to all pupils regardless of race or culture. All facets of education have been affected, in particular evaluation at all three levels of education, namely primary, secondary and tertiary levels. Due to the changed circumstances and new approaches in the teaching and instruction process in South Africa (Steyn, 1989:448-459), a re-evaluation of our existing evaluation methods has become imperative.

Because of the central role of evaluation in the educational process and the far-reaching consequences it holds for the future of learners (e.g. pass/fail, having to repeat a full year, the effects on learner achievement, etc. (Van Aardweg & Van Aardweg, 1988:8), it is a prerequisite that evaluation methods be of a very high standard and quality. The instruments used to gather the necessary information about learners’ progress and achievement in particular must be valid, reliable and relevant enough for that purpose (Cangelosi, 1990:27-36).
It is primarily because of a quest for this high standard and quality of the evaluation methods, and in particular the measurement instruments at our tertiary educational institutions, that this research is undertaken. There have long been voices criticizing the existing measurement instruments which renders such a research project even more urgent and essential (Garcia & Pearson 1994:337-335; Cangelosi, 1990:3; Cawood, Strydom & Van Loggerenberg, 1988:126-128; White, 1983:131-133; Frith & Macintosh, 1988:79; Bergman, 1981:128, etc.).

Although the focus in this research is on one particular subject and one particular institution only, which was necessitated by a number of factors, the overall intention with this research is that all educational institutions in South Africa and even beyond its borders, will re-examine their existing evaluation and measurement methods with the intention of upgrading and improving them.

With this introduction as a means of orientating the reader, a more systematic exposition of this research will be supplied below.

1.2 BECOMING AWARE OF THE PROBLEM

It was while the researcher was busy with a work-related project on summative evaluation of learner achievement at Technikon SA, that an awareness of a number of shortcomings in the existing evaluation methods started to emerge (Van Koller, 1994:55). This particular research project focused specifically on the first component of evaluation, that is measurement and the instruments used in the measurement process (which will be fully described in section 1.8.3).

During the course of this project, it was determined that, despite a number of decided advantages, the existing measurement instruments (primarily the written test and examination) used at most educational institutions in South Africa, have a number of serious shortcomings which include the following:
Written forms of measurement are focused mainly on the evaluation of the cognitive skills of learners at the expense of the affective and psychomotor skills.

The written examination can be a major disadvantage for learners who are not as capable as others in expressing their knowledge, comprehension and skills by means of the written word.

The extreme conditions of stress and tension normally accompanying the process of the written examination, in many cases prevent learners from effectively answering questions and performing optimally.

The inconsistency arises when learners, who do not really have a grasp of the total curriculum, are able to pass the examination merely on the basis of "spotting" (guessing) the correct questions in advance, while learners with a better understanding of the work fail because of not preparing the right questions.

In relation to the multilingual context in South Africa, the written examination appears to benefit learners from certain language groups at the expense of others.

In relation to the multicultural context of South Africa, it also seems as if the written examination is more biased towards learners from a predominantly western or European cultural background at the expense of learners from an African culture (Van Koller, 1994: 22-25; 29-31; 34; 37, 40 - 45).

These are only a few examples of the numerous flaws in the existing measurement instruments. These, coupled with the fact that numerous authors have complained about these shortcomings, make the need for the examining of alternative, more suitable, valid and adequate measurement instruments imperative (Garcia & Pearson, 1994: 337 - 355; Steyn, 1989: 448 - 459; Sentson, 1994: 109-113; Cangelosi, 1990: 3, NEPI, 1992: 21-33; Cawood et al, 1988: 126-128; White, 1983: 131-133; Frith & Macintosh, 1988: 79,80; & Bergman, 1981: 83, 128).
In order to supply a more precise exposition of the exact difficulties of the existing measurement instruments at Technikon SA, the next section will be devoted to a clear formulation of the problem.

1.3 STATING THE PROBLEM

In order to enable the reader to grasp the extent of the problem, it is necessary to supply the full context.

First, there are the numerous shortcomings in the existing measurement instruments as pointed out by the various authors cited in point 1.2. The final two shortcomings mentioned in 1.2, namely the multilingual and multicultural contexts of South Africa in which a large percentage of the learner population receive instruction by means of second or third language (Sentson, 1994: 109-113), probably form the most serious obstacles in the existing measurement instruments. Garcia & Pearson (1994: 340-355), point out the severe discrepancies caused by the existing written measurement instruments in the United States of America’s multilingual and multicultural society.

Second, there is the unique nature of Technikon SA that raises a further question about the suitability and adequacy of the existing measurement instruments. There are a number of dimensions relatively unique to Technikon SA that are not adequately taken into account in the use of the existing measurement instruments, because these measurement instruments were designed originally for use in residential (face to face) educational institutions. These particular dimensions at Technikon SA are the following:

- The aspect of distance learning which involves the absence of face-to-face contact and interaction between lecturer and learner.

- The aspect of self-instruction and own pace of study by learners. The course material is the primary source of instruction.
• The practical and technically oriented nature of Technikon SA courses. The method of transferring this practical and technical instruction to learners by means of the written word requires special skills, as would also the evaluation methods used.

There are a few other aspects that are not unique to TSA only but that require special attention insofar evaluation methods are concerned. Such aspects are for example:

• The aim of the mastery of course material and skills.

• The aspect of second and/or third language learners.

• The aspect of multiculturalism.

Third, the aspect of new approaches to the total teaching/instructing process in South Africa (Steyn, 1989: 448 - 459; Van Koller, 1994: 51-54) places further demands on existing evaluation methods and measurement instruments. The move away from the primarily product-based teaching (knowledge and content) to a more process- or performance-based form of instruction (Eisner, 1986:71-84; Rowntree, 1990: 312; Dahlöff, Harris, Shattock, Staropoli & In’t Veld, 1991: 119, 120) will most probably create a need for measurement instruments that are in line with or congruent to this new approach.

Against the context of these three areas of concern, the following question undoubtedly arises:

Can the written form of measurement instruments still be regarded as suitable, adequate and acceptable enough for the purposes of summative evaluation of learner achievement in Business Management at Technikon SA?
In order to answer the main question, a number of related or sub-questions will need to be addressed. These sub-questions are the following:

- First, what is the exact role and function of evaluation and measurement in the total educational process? It is essential to answer this question, because it will help the reader to gain a clear understanding about the role of evaluation and measurement in education as well as the importance of this research not only for Technikon SA, but also for other educational institutions and employment organisations that might be affected.

- Second, what are the shortcomings or disadvantages of the existing summative evaluation methods that were identified in literature. A description of such shortcomings will help to create a better understanding of the causes of this research. To answer this question will require a thorough literature review in order to provide a balanced view on the existing evaluation methods.

- Third, what alternative measurement instruments are available and which alternative(s) will best suit Business Management? This question will require extensive research into existing alternative measurement instruments and a thorough scrutiny of these alternatives against the backdrop of certain criteria.

- Fourth, does any correlation exist between the curriculum objectives of BM and the measurement instruments used in BM? The answer to this question will help the researcher to determine whether the measurement instruments used in BM are suitable in terms of the BM curriculum objectives. This question will require an analysis of both the BM curriculum objectives and the measurement instruments used in BM and a comparison between them.

- In the final instance, what are the views of learners in BM at TSA regarding the existing measurement instruments used in BM? This question needs to be answered, because it is essential to determine whether the stakeholders, in this instance the learners, indeed feel dissatisfied with the existing measurement
instruments and whether they are receptive to alternative instruments, or whether they are still content with the existing instruments.

The purpose of this research is to answer the above questions and to produce valid and scientifically-based data. It is, however, necessary to supply a more precise outline of the various aims of this research which will be done in the next section.

1.4 AIMS AND PURPOSES OF THIS RESEARCH

The overall aim of this research is to examine the possible implementation and use of alternative measurement instruments for the purpose of summative evaluation of learner achievement at Technikon SA as a whole, as well as at all other tertiary educational institutions in South Africa. However, the unique nature and requirements of different subject disciplines, different educational institutions, and the immense task of examining all of them on a single basis, render it impossible to be broadly prescriptive in this research.

However, it is hoped that this research will initiate similar projects by scholars in other subject disciplines at Technikon SA and at other tertiary institutions. It is also hoped that the findings in this research, the conclusions and the resultant recommendations, will be of value to similar educational institutions and other affected bodies, such as employment organisations and even the Small Business Development Corporation of South Africa.

The more specific aims of this research are as follows:

- First, to give the reader and fellow scholar a very clear understanding of the role and function of evaluation in general and of measurement and measurement instruments in particular in the education process.

- Second, to examine whether the existing measurement instruments used in Business Management at Technikon SA, are still widely accepted by the
stakeholders involved, or whether there is indeed a need for alternative measurement instruments.

- Third, to launch an extensive research into various existing alternative measurement instruments and, by means of a set list of requirements, to select those alternatives best suited for Business Management at Technikon SA.

- Finally, to make recommendations based on the findings in this research, that will be useful not only for Business Management at Technikon SA, but also for various other subjects at Technikon SA and at other educational institutions in South Africa.

With the central aims of this research stipulated, it would be suitable at this stage to supply an indication of the process through which the researcher intends to achieve these aims. In the following section the particular methodology used in this research, is described.

1.5 RESEARCH METHODOLOGY

The methodology decided upon for the purposes of this research, consisted of two main methods, namely a thorough and comprehensive literature review of the focus area of this research, and different qualitative research methods.

1.5.1 Literature review

It is necessary to substantiate why these methods were selected. Starting with the literature review, a number of considerations can be mentioned of which the following probably form the crux:

- First, any form of scientific inquiry requires a profound theoretical foundation in order to make such inquiry logical, valid and scientifically accountable (Mulusa 1990:115).
• Second, it is always necessary and, in fact, essential to determine what previous research had already been undertaken in the said field of inquiry and what knowledge and information were available. This would obviate new research being a duplication of existing data, but to be a contribution to the knowledge and information in that field.

Eichelberger (1989:70) supplies the following compact summary of the main purposes of a literature review:

• To learn the history of the problem
• To become familiar with the theoretical background of the problem
• To assess the strengths and weaknesses of previous studies
• To identify viable ways to study the problem
• To develop a conceptional framework and rationale for the present study

These are indeed the purposes of the literature review in this research. It is also important to mention that the references and bibliography comply with certain basic standards and requirements.

• First, in terms of recency of sources, an effort was made to consult the most recent literature relating to the particular topics of this research. It was only in exceptional cases, i.e. where more recent sources dealing with a particular aspect were not available or where primary sources in certain instances had to be used, that sources older than 1980 were consulted.
• Second, an attempt was made to use mainly primary sources, but because of language and availability considerations, secondary sources were used in certain circumstances.
• Third, an effort was also made to use mainly sources that related specifically to the particular aspect or topic on hand in order to obtain the most comprehensive reports on these topics.
In conclusion, it would be fair to mention that the different sources consulted in this research, did not all relate to the fields of didactics or education as such, but to other subject disciplines, for example Social Work, Psychology, Human Resources and so forth.

1.5.2. The qualitative research method

In order to substantiate the choice of the qualitative method of research for the purposes of this research, three questions needed to be answered:

- What is qualitative research?
- Why qualitative research?
- How does this research method operate?

1.5.2.1 What is qualitative research?

Qualitative research is described as follows by a number of authors:

Strauss & Corbin (1990:17) define qualitative research as any kind of research in which findings are not arrived at by means of statistical procedures or other means of quantification.

Tesh (1993: 3, 55) argues that qualitative data implies any data that is not regarded as quantitative, in other words, data that is not expressed by way of numbers.

Sherman, Webb & Andrews (1984:509) reason that the term “qualitative” implies a direct concern with experience felt, lived or undergone by human beings. They conclude that the main aim of qualitative research is to understand experiences as closely as possible to the way in which the particular person(s) experienced it.

Some authors regard the qualitative approach to research as the complete opposite of the quantitative approach. The two main differences between qualitative and
quantitative research appears to be the method of data collection and the overall or primary aims of each approach.

The assembling of data in qualitative research proceeds mainly by means of personal interviews, questions and observation (Tesh, 1993: 2; Strauss & Corbin 1990: 18; Marshall & Rossman, 1995: 78; Jacob, 1988: 184). This form of data collection puts it into sharp contrast with the quantitative method which relies strongly upon statistical procedures (Strauss & Corbin, 1990:17).

Lemmer (1989: 7, 8) on the other hand states that the overall purpose of quantitative research is to explain, whereas the primary aim of qualitative research is to understand the life-world of individuals or groups.

Despite the fundamental differences between the two research approaches, certain authors seem to believe that these two research approaches can be used on a complementary basis in the same research project (Strauss & Corbin, 1990:19 and Laney, 1993:22). Laney (1993:12) has pointed out on a number of occasions that qualitative researchers have utilized various quantitative techniques for example, experiments and attitude surveys near the end of their studies.

In summary the question: What is qualitative research? can be answered in the following way: Qualitative research is a form of research in which the findings are not based on quantitative techniques, but on personal, first-hand interaction and observation in order to understand the feelings and experiences of people and give a valid and true report thereof.

1.5.2.2 Why qualitative research?

The considerations which motivated the researcher to choose this method, will be supplied in random order.
• First, the growing popularity of qualitative research in the Social Sciences especially in the 1980s, played a role in this decision. Tesh (1993: 2) and Laney (1993: 23, 24) mentioned the rising interest that researchers over a wide spectrum, even well-known quantitative researchers, were displaying in qualitative research because of its closeness to natural and real-life settings.

• A second factor that played a coercive role, is the aspect of suitability. Strauss & Corbin (1990:19) state for example that qualitative research is probably the most effective method for the study of organisations, groups and individuals. Tesh (1993:2) also underscores this view. Since this particular research will focus on groups of people as well as individuals, the qualitative method of research is most suitable.

• A third, and probably the strongest motivating factor, was the aspect of validity. Authors such as Laney (1993: 24 -26), Strauss & Corbin (1990:19) and Taylor & Bogdan (1984: 6 - 15), argue that the findings in qualitative research are more valid than those of quantitative research if the research focuses on human beings. These scholars supply the following reasons for this contention:

  - Human behaviour is greatly influenced by the particular setting or context in which it takes place. Research must therefore be undertaken in settings where all the relevant variables are present and operating. In quantitative research this is not possible because the setting is foreign in the sense that it does not represent a microcosm of the real-world setting and it is artificial. In contrast, in qualitative research the primary intention is precisely to capture real-world and real-life situations by the researcher becoming part of that setting (Laney, 1993: 24 - 26).

  - Experimental research does not enable the researcher to understand the framework within which people interpret their thoughts, feelings and actions. However, qualitative research which compels the researcher to immerse himself into the everyday life of the setting(s); to interact with the participants and
value their personal views and perspectives on various issues, make it possible
for the researcher to achieve this aim of understanding (Taylor & Bogdan,
1984: 6, 15).

- Qualitative methods are effective in uncovering and understanding what lies
behind a phenomenon that is not so well-known. It can therefore supply the
deepest and most precise details of phenomena that cannot be conveyed easily
by means of quantitative methods (Strauss & Corbin 1990:19).

- Quantitative research sometimes oversimplifies certain facts or data which may
in fact be very complex. Lancy (1993: 25) illustrates, for example, that
procedures relating to the different models and designs used in quantitative
research require variables that can be scaled, that is numerically scored or
rated. The qualitative researcher, however, would insist that not all of the
reality that forms part of education is reducible to variables.

- Many recent debates on research methods suggest that quantitative research in
many instances obscures the very phenomenon under investigation, while this is
not the case as far as qualitative research is concerned (Lancy, 1993: 26).

This list is not complete, but it is evident from the reasons supplied, that qualitative
research stands in strong contention with quantitative research, which was previously
accepted as the only valid method of research, in terms of validity. The
arguments presented by the above-mentioned scholars as well as some others (Tesh,
1993: 2, 3, 55 and Jacob, 1988: 184, 185) were convincing and strongly influenced
the researcher’s choice of research method.

1.5.2.3 How does qualitative research operate?

Finally, the question on how this method operates will be answered by a brief
description of the process of qualitative research. As mentioned previously, the main
forms of data collection are the personal interview and observation which can be
direct or indirect. After collecting data, which can be by means of audio tape recordings, the data must be analysed and interpreted by the researcher, after which the process of written and/or verbal reports of the data and events is completed (Strauss & Corbin, 1990: 18-20).

Particular characteristics and principles are involved in the process of qualitative research. Jacob (1988: 184), for example states that qualitative research is normally seen as free from pre-determined theories and questions. Questions and theories only emerge after the data collection process, not before. Furthermore, authors such as Sherman, Webb & Andrews (1984: 508) also support this view. This view also accounts for the absence of a preconceived hypothesis in this research. There is also the quest to operate in the most natural setting conditions as possible to avoid artificiality. The total process of qualitative research is also a very transparent one (Lancy, 1993: 26, 27).

In conclusion it is also essential to note that qualitative research is not a mere uniform type of research. There are a variety of qualitative research methods described by different authors. Some authors apply different names to the different methods of qualitative research, but the basic methods reveal a common basis.

Tesh (1993: 43-49) supplies the following list of qualitative research methods:

- Naturalistic inquiry
- Illuminative evaluation
- Transcendental realism
- Ethnography
- Ethnography of communication
- Phenomenological studies
- Educational connoisseurship and criticism
- Critical educational science with as goal action research as empowerment
- Collaborative research
- Phenomenography
Jacob (1988 : 185-187) lists the following domains of qualitative research:

- Human ethology
- Ecological psychology
- Holistic ethnography
- Cognitive anthropology
- Ethnography of communication
- Symbolic interactionism

Strauss & Corbin (1990:20) distinguish the following methods:

- Grounded-theory (of which Strauss was a pioneer);
- Ethnography
- Phenomenological approach
- Life histories
- Conversational analysis

Peshkin (1993: 500 - 504) groups these different qualitative methods into certain categories, which indicate the main emphasis, namely:

- Description
- Interpretation
- Verification and
- Evaluation.

For this research it was decided to use the focus group interview, which is a particular type of interview, as the main qualitative method. In addition to the focus group interview two supplementary methods, the survey and document review, were also used. However, more detail about the specific methods used in this research will be provided in the chapters that deal specifically with the use of those research methods, that is chapters 5 and 6.
With this broad exposition of the main research methods used in this research, the exact demarcations of this research and its limitations will be discussed next.

1.6 DEMARCATIONS AND LIMITATIONS

Because of the immense scope of the key topic of this research, that is evaluation, it was necessary to supply an exact demarcation of this research for the purposes of clarity and to point out its limitations.

- First, this research focuses on one particular component of evaluation only. As will be explained in 1.8.1, the concept of evaluation consists of three distinctive components, namely measurement, assessment and judgement. This research focuses only on the first component, measurement, more specifically, the various measurement instruments used to collect information or data for the further processes of evaluation. There will, however, be a clear definition of the other two components of the concept of evaluation in order to indicate their differences and their relationship to the first component.

- Second, the measurement instruments examined in this research, were the existing ones used at Technikon SA, with perhaps one exception, for the collection of information on learner achievement in a particular subject, as will be discussed in chapter 3.

- Third, the particular subject focused on was Business Management. This includes Business Management I, II, and III. Although the overall intention with this research was to make recommendations on a broad subject disciplinary basis, the recommendations at the end of this research can only be regarded as directly applicable to Business Management I, II, III at Technikon SA.

- Fourth, this research was concerned with summative evaluation as opposed to formative evaluation. Summative evaluation, as clearly defined in 1.8.2, refers to a more final type of evaluation normally conducted towards the end of a course or
study period, where as formative evaluation refers to a more continuous process of evaluation. References to formative evaluation were however made in the course of the research.

- Finally, the focus in this research was also on one specific educational institution, Technikon SA. This research therefore took specific cognisance of the unique nature of Technikon SA as a technical, distance education institution. All the findings and recommendations proposed in this research will therefore only apply to Technikon SA, although it is hoped that other educational institutions might implement some of these recommendations, where applicable.

In this brief exposition of the boundaries and limitations of this research, it would also be necessary to contextualize the selection of the subject “Business Management” for the purposes of this research.

1.7 WHY BUSINESS MANAGEMENT?

The reasons why the subject “Business Management” was selected for this research are not based on any specific professional, scientific or fundamental reasons. The researcher for example, is not a lecturer in the subject, nor does he have any tertiary qualifications in the subject. In fact, the researcher specialises in a totally different field than Business Management, that is Didactics. Why then Business Management?

- First, because of the extensive scope of the concept of evaluation and the numerous subjects offered at TSA, the researcher in consultation with his promoter decided it wise to focus on one subject only.

- Second, it is important to state that for the purposes of this research, almost any subject that uses the written examination for summative evaluation purposes, would have qualified for selection. The main purpose of this research was to establish how adequate and acceptable the written examination as such still is at TSA.
• During the beginning phases of this research the researcher, due to the nature of his work, worked in close co-operation with the programme group (department) Business Management. There was therefore regular interaction and consultation between lecturers from Business Management and the researcher.

• The subject “Business Management” can also be regarded as a typical TSA subject, because there is normally a high learner enrolment in the subject and the subject is not restricted to one career path only. Students from a number of different fields can enrol for the subject.

• Business Management also posed an interesting challenge to the researcher because of the fact that partly alternative evaluation methods (measurement instruments) are already used in the subject. Both the case study approach and the open-book examination can be regarded as a move away from the traditional or conventional written examination (see section 3.3.6 and 4.5.2.4.4 for more elaboration). However, both methods are still based on the written word and are also subject to some of the main limitations of the written examination (see 3.5.1.2).

• A final, yet minor consideration, was the fact that the researcher received instruction in the subject until matric level and could therefore identify with the contents and nature thereof.

In conclusion it is necessary to state that the researcher is aware of the fact that, because he is not a specialist in the subject area, he might not be aware of certain unique features of the subject. The researcher therefore accepts the fact that any recommendations or statements made in this research, are subject to more specialist knowledge from scholars and professionals in Business Management. The researcher does in particular not intend to be prescriptive with any recommendations. The main intention is merely to stimulate a greater sensitivity among all academics at TSA and elsewhere concerning the need for suitable and adequate forms of evaluation.
For interest sake, it should be noted that the original name of this subject was Business Economics. Since the beginning of 1997 the name has officially changed at TSA and elsewhere to Business Management.

The next section will be devoted to a definition of certain central concepts that played significant and dominant roles in this research.

1.8 CONCEPT DEFINITIONS

The central role of the following concepts makes it essential that the reader understands their exact meanings and uses in this research. These concepts are the following:

- Evaluation
- Summative evaluation
- Measurement
- Measurement instruments
- Alternative measurement instruments
- Learner
- Achievement

1.8.1. Evaluation

To locate an all-inclusive and universally-accepted definition of the concept of evaluation, is difficult because of its immense scope and the differing opinions of a wide spectrum of evaluators (Thorpe, 1988: 5). Very few authors and specialists in the field of educational evaluation supply a comprehensive and inclusive definition of the concept. Normally they only emphasize one or, at the most, two components of evaluation in their definitions as will be indicated below.
• Stufflebeam (as quoted by Mehrens & Lehmann, 1991: 4): “Evaluation is the process of delineating, obtaining, and providing useful information for judging decision alternatives.”

• Jacobs (1986: 35,36): “Evaluering is 'n meer komplekse begrip en sluit eintlik meting in” (Evaluation is a more complex concept and in fact includes measurement (own translation)).

• Duminy & Söhngé (1987: 113): “The information gathered by measurement is analyzed and interpreted to find out how successful the teaching has been. This is called evaluation.”

• White (1983: 122) “It (evaluation) involves the interpretation of marks, symbols or rankings, and of other measurements, leading to decisions about subsequent action”. (Own inclusion).

• Hopkins & Antes (1990: 542): “Evaluation (is) the continuous inspection of all available information concerning the student, teacher, educational program and the teaching-learning process to ascertain the degree of change in students and to form valid judgements about the students and the effectiveness of the program.” (Own inclusion).

• Lawton & Gordon (1993: 82): “Evaluation (is) a term which may either refer to the general process of judging the worth of an educational programme, including judgements about the quality of its content, or more specifically to measurements of the effectiveness of learning experiences.” (Own inclusion.)

• Barrow & Milburn (1990: “…many subscribe to the principle that the job of evaluation is to make an overall judgement: to assess the educational worth of an activity or programme.”

• Cangelosi (1990: 25): “An evaluation is a value judgement.”
A close scrutiny of these different definitions of evaluation results in the conclusion that the concept of evaluation consists of three distinct components. These components are:

- First, the collection of information or data, as indicated by the first three definitions. This process of collecting data is generally known as measurement and is explained fully in 1.8.3. Thus it seems as if measurement is the first component of evaluation.

- Second, evaluation consists of a process of analysing and interpreting that data assembled by means of measurement, as indicated by the next two authors. In this research, the processes of analysis and interpretation are regarded as two sides of the same coin, because the one (interpretation) is a logical result/consequence of the other (analysis). These processes of analysis and interpretation can be seen as assessment. Although the term “assessment” is frequently used as a synonym for evaluation itself (Barrow & Milburn, 1990: 123), it seems that the term can definitely be linked to these processes. Rowntree (as quoted by Frith & Macintosh, 1988: 4) for instance defines the process of interpretation of information as assessment. Barnhart & Barnhart (1982: 123) define the word, assess, to mean: “examine critically and estimate the merit, significance or value of”. This is exactly what takes place in the processes of analysis and interpretation. Hopkins & Antes (1990:2) and Lawton & Gordon (1993: 82) regard assessment as one of the components of evaluation and not the full process of evaluation itself. In conclusion, assessment, which refers to the analysis and interpretation of data, seems to be the second component of evaluation.

- Third, evaluation consists of the formulation and issuing of a judgement, as indicated by the last three definitions. This judgement is a result of and based on the two aforegoing components of evaluation. In other words, judgement is not possible if both the two components, measurement and assessment, have not first
been employed. It seems therefore that the issuing of a judgement is the third component of evaluation.

It is therefore very clear that the three components of evaluation mentioned, form an integral part of the total process of evaluation. An inclusive definition should therefore refer to all three these components. A definition that comes the closest to this aim, is the one supplied by Thorpe (1988: 5), namely:

- "Evaluation is the collection, analysis and interpretation of information about any aspect of a programme of education and training, as part of a recognised process of judging its effectiveness, its efficiency and any other outcome it may have." (Own highlighting).

A more concise version of this definition would be as follows:

- (Educational) evaluation is the collection of information or data about any educationally-linked item after which such information or data is analyzed and interpreted in order to reach and issue a value judgement concerning that particular item, or part of it.

The above exposition of the process of evaluation is also underscored by Sims (1992: 4) which lends even greater emphasis to this contention.

For the purpose of this research therefore, this definition of evaluation will be the accepted one. See also the latter part of section 2.2.6, chapter 2, for further elaboration on the use of the term “evaluation” instead of “assessment”.

With this as a point of departure, the other concepts will be defined.
1.8.2 Summative evaluation

Summative evaluation is concerned with the evaluation of the final outcome of a programme or period of study. Summative evaluation is normally seen as the direct opposite of formative evaluation, which refers to a more continuous process of evaluation.

Formative evaluation is described by Thorpe (1988: 9,10), Rowntree (1990: 302, 33) & Bergman (1981:54) as a process of evaluating the progress made towards the achievement of programme goals during the implementation of those goals. The aim is therefore, not to measure the effectiveness of a programme, but to identify any changes required to improve it and make it more likely to achieve its goals in the long run.

Summative evaluation, on the other hand, is defined by Hopkins & Antes (1990 : 551) as: “A terminal evaluation employed in the general assessment of the degree to which the larger outcomes have been obtained over a substantial part of or all of a course. It is used in determining whether or not the learner has achieved the ultimate objectives for instruction which were set up in advance of the instruction.”

The emphasis in this research is in line with this definition of Hopkins & Antes. The summative evaluation of learner achievement in Business Management I, II, and III, will therefore be the focal point. The third concept will be defined next.

1.8.3 Measurement

Measurement, as touched upon in 1.8.1, refers to the first component of evaluation which entails the process of collecting information or data. As pointed out earlier, measurement is regarded as an essential component without which the full process of evaluation is impossible. For the purposes of this research and for an exact definition of the concept of measurement instrument, it is necessary to supply an exact and comprehensive definition of the concept of educational measurement.
Duminy & Söhge (1987: 113) define measurement as follows: "Measurement has to do with the collection of data or information and refers to the use of tests and examinations in measuring the changes brought about by education."

Van Aardweg & Van Aardweg (1988: 145) define it as follows: "In education, measurement refers to the use of objective tests to yield quantitative data".

Mehrens & Lehmann (1991: 4) supply the following definition: "Using observations, rating scales, or any other device that allows us to obtain information in a quantitative form, is measurement."

Cangelosi (1990: 25) defines measurement as: "The process by which facts or data are gathered through empirical observation."

Barrow & Milburn (1990: 123) describe measurement in the following way: "Measurement generally has a very precise denotation: assigning numerals to objects or events according to rules. Consequently, those interested in measurement typically search for observable indications of properties of objects or events under study, apply rules to those properties, and then allocate a number to the result, often using statistical techniques."

Hopkins & Antes (1990:7) define it as: "A process that assigns by rule a numerical description to observation of some attribute of an object, person, or event."

After analysing these definitions, it would appear that the concept of measurement consists of four specific elements, namely:

- the gathering/collection of information or data;
- observation, rating scales or other processes;
• assigning of numerals (quantities) to a particular attribute of an object, person or event taking place; and
• based on specific, predetermined rules.

When the term "measurement" is used in this research, it will refer to the four elements specified above. The researcher also wants to make it clear that the term measurement is used in this research to indicate the exact section of the evaluation process that is of concern in this situation. Second, it is also important to note, that unless the total system of evaluation is changed from the awarding of numbers to indicate the achievements of learners, to a system where all progress and achievement are indicated by means of grades, the term "measurement" can not really be abandoned. As long as numbers are assigned to indicate learner achievement the process of measurement is operating. However, the researcher is not "married" to this term, but for the purposes of this research, it was regarded the most suitable term to use. These comments were made for the sake of the argument that the term "measurement" should no longer be used, rather "assessment".

In conclusion, it is also very important to note that measurement pertains to the properties of objects and not to those objects themselves (Sirotnik, 1987: 23; Hopkins & Antes, 1990: 8). This implies that in the context of this research, it is not the measurement of the learners themselves that will be examined, but the measurement of their academic achievement in Business Management I, II, and III.

A definition of measurement instruments will be expounded next.

1.8.4 Measurement instruments

In the light of the exposition of the meaning and elements of the concept of measurement as supplied in 1.8.3, the concept of measurement instrument will refer to the second element of measurement, namely the means or process by which the particular information or data will be collected.
The following authors define the concept of measurement instruments as follows:

- Barrow & Milburn (1990: 123): "Testing and examining usually suggest the application of a standardised (as in tests of basic skills) or teacher-devised instrument to gauge the progress of learners during or at the end of a course of study."

- Hopkins & Antes (1990:7). "In general, an observation made by a measurement process involves a measurement instrument. Rather than making a judgement about an observation, the observer uses an instrument to quantify the attribute of concern. The instrument is used because of the objectivity provided by the properties of the instrument."

- Jacobs (1986: 35). "Toetsing en eksaminering kan geklassifiseer word as meetinstrumente". (Testing and examination can be classified as measuring instruments - own translation.)

- Cangelosi (1990: 25): "Testing are planned measurements by which teachers attempt to create opportunities for learners to display their achievement, relative to specific goals."

- Van Aardweg & Van Aardweg (1988: 145): "The measurement instrument (the test) allows for the quantification of a learner’s performance."

These definitions tell us the following about measurement instruments: First, in order to measure anything, a measurement instrument is required. Second, the data collected by the measurement instrument forms the basis for further analysis and interpretation. Third, the main types of measurement instruments used in educational evaluation are tests and examinations.
Finally, measurement instruments have an additional purpose (apart from the collection of data) that is to quantify the result of the observation(s) by ascribing numerals to it.

With the meaning of the concept of measurement instrument already clearly explained, an explanation of what is meant by alternative measurement instruments will be given.

1.8.5 Alternative measurement instruments

Alternative measurement instruments in education imply types of measurement instruments that differ from the two conventional and dominant ones mentioned in 1.8.4, namely tests and examinations. Garcia & Pearson (1994: 356) state one view which regards alternative evaluation as "a description intended to capture the sense of rejecting the conventions of formal assessment." The emphasis on alternative evaluation in this research is not so much on the rejection of the conventional measurement instruments, but rather on searching for more effective instruments.

A number of different alternative measurement instruments already exist. Frith & Macintosh (1988: 52) supply a summary of a number of such alternatives, namely:

- Oral assessment
- Aural assessment
- Coursework

Bergman (1981: 183 - 240) also discusses a few alternatives, namely:

- Grading
- Rating
- Social measurement
These examples are only basic categories of alternative measurement instruments, because in each case a variety of alternatives are involved. Garcia & Pearson (1994: 356) also supply a list of alternative measurement categories, namely:

- Performance assessment
- Authentic assessment
- Portfolio assessment
- Situated (or contextualized) assessment
- Dynamic assessment
- Curriculum-embedded assessment
- Assessment by exhibition

These lists of possible alternative measurement instruments are not exhaustive. They merely serve to supply examples of what is meant by alternative measurement instruments in this research. Because a particular chapter in this research (chapter 4) is devoted specifically to an examination of these various alternative measurement instruments, no description or further elaboration on these alternatives will be supplied at this stage.

The next concept to be defined in this chapter, is the concept of learner.

1.8.6 Learner

The term “learner” is an inclusive term used at Technikon SA to refer to learners of any age. The term “pupil” normally refers to a learner still at school level (primary and secondary), while the term “student” normally refers to the learner at tertiary level. It has, however, been noted that the tertiary learner can range from a young age (17, 18 years of age) and fresh from school to senior adults who had been breadwinners of families for some time. Even the majority of lecturers at tertiary institutions are also normally in the process of upgrading or furthering their studies. This is of course also the case at Technikon SA and specifically in the field of Business Management. The question therefore arose whether the term “student” was adequate to represent such a
wide and complex group. As a result of this process of re-examining this term, a decision was taken to use the term “learner” instead. This term appears to be most inclusive because it would include every person, irrespective of age, race, gender, status or position, involved in the process of learning.

The last concept that will require specific definition, is the concept of achievement.

1.8.7. Achievement

Achievement (or learner achievement) refers specifically to what the learners have succeeded in accomplishing in terms of the objectives of a particular course and in terms of the criteria that will determine whether they have passed or failed that course.

Two possible meanings of the word “achieve” can be:

- “to carry out to a successful end”
- “to accomplish” (Barnhart & Barnhart, 1982: 79)

The word “achievement” means: “something achieved or won by exertion” (Barnhart & Barnhart, 1982: 79). In order for learners to pass a certain course, they must exert themselves by studying and mastering certain course contents or skills.

Van Aardweg & Van Aardweg (1988: 8) state that achievement is almost always linked to standards recognised by others. These standards in terms of Technikon SA would refer to the particular standards or criteria which determine whether a learner has accomplished certain specified requirements. These requirements determine whether learners qualify to be awarded a certain credit in the form of a certain symbol or on the basis of a pass/fail report. At present, the requirements are that learners should achieve a minimum of 40% in the examination, but an overall average of 50% (which is constituted from both examination and yearmark on a 60%/40% basis) in order to pass a particular course.
At Technikon SA, a technical and practically-inclined institution, the focus of learner achievement should not only be on obtaining of cognitive and academic knowledge and skills, but also on the achievement or mastery of particular practical, psychomotor and task-oriented skills. These skills must enable learners to perform the tasks and duties for which they were trained. However, whether this is indeed the case will be determined by the specific curriculum or course objectives. This aspect will be examined in the course of this research, specifically in relation to Business Management I, II and III.

These comprehensive definitions of the most central concepts of this research, should contribute to a greater clarity and a better understanding by the reader. At this stage, a concise exposition of the main structure and the further development of this research will be discussed.

1.9 STRUCTURE OF THE RESEARCH PROGRAMME

The further structure of this research will follow the arrangement of the different sub-questions posed in 1.3 very closely. The reason for this pattern is so that each of the subsequent chapters will attempt to answer those questions.

Chapter two will focus mainly on the question: What is the exact role and function of evaluation and measurement in the total education process? The answer to this question will require an extensive theoretical foundation of the phenomena of evaluation and measurement in order to portray the central and essential role they play in the educational process. Such a theoretical foundation will also serve to further substantiate the importance and value of this research. This chapter involved mainly a comprehensive literature review as main research method.

Chapter three will concentrate on an examination of the existing measurement instruments used at TSA and in particular in Business Management. The objectives of this examination will be to determine if there are indeed any shortcomings or
disadvantages attached to the existing measurement instruments and what such disadvantages in fact entail. To achieve these objectives will require the following:

- First, a description of the existing summative measurement instruments at Technikon SA in general and in Business Management in particular.
- Second, a precise portrayal of the nature and extent of the shortcomings thereof as indicated by authors in the field.

Chapter four will attempt to answer the question: What different alternative measurement instruments are currently available and how feasible are they for use at Technikon SA, in particular for Business Management? This will entail a description of various alternative measurement instruments and a scrutiny of the advantages and disadvantages of such instruments. This process required mainly a literature review.

Chapter five will attempt to determine whether there is a positive correlation between Business Management curriculum objectives and the measurement instruments used in BM. The aim is to establish whether the measurement instruments used in BM are adequate enough to measure to what extent the curriculum objectives were realised.

Chapter six will focus on the views and perceptions of one of Technikon SA's main stakeholders, the learners, concerning the existing measurement instruments used in Business Management. By means of the qualitative research methodology, and more specifically, the focus group interview, an attempt will be launched to capture their views and the requirements they would like to set regarding preferred measurement instruments.

Chapter seven will conclude this research and will consist of a summary of the entire study, a summary of the main findings, a number of conclusions drawn from these findings, a number of recommendations based on these conclusions and finally, indications of further possible areas of research relating to the theme of this research.
With this brief portrayal of the further structure of this research, it is hoped that a clear indication of the direction, plan and programme of this research has been supplied. In conclusion the researcher will provide a brief summary of this chapter.

1.10 SUMMARY

In this chapter an introduction to the main theme of this research was supplied, that is the possible need for alternative measurement instruments in summative evaluation of learner achievements in Business Management at Technikon SA. Following this, a description of how the researcher became aware of an existing problem (namely the existence of numerous shortcomings) in the existing evaluation methods at Technikon SA. Then, the exact problem that needed to be addressed, was stated, that is whether the existing measurement instruments for Business Management at Technikon SA could still be regarded as suitable, adequate and acceptable enough or whether they had to be replaced by alternative measurement instruments. After this, the specific aims and purposes of this research were presented. Next a clear exposition of the research methodology used in this research, namely a comprehensive literature review and the qualitative research approach was given. The exact demarcations and limitations of this research were then explained followed by an explanation why Business Management was chosen as a suitable subject to conduct this research in. This was followed by a clear definition of the most central concepts used in this research. These central concepts are: evaluation, summative evaluation, measurement, measurement instruments, alternative measurement instruments, learners and achievement. Finally, a concise description of the further structure and programme of this research was supplied.

The following chapter will be devoted to a theoretical foundation of the role and function of evaluation and measurement in the total educational process.
CHAPTER 2

A BROAD OVERVIEW OF THE PROCESSES OF EVALUATION AND MEASUREMENT

2.1 INTRODUCTION

In chapter 1 the underlying problem that formed the basis of this research, was described. Also, the purposes of this research; the research methodology employed to collect the required information; the demarcations and limitations thereof, a definition of the most central concepts that frequently appear; and the basic structure were provided.

This chapter is intended to provide a broad overview of the processes of evaluation and measurement so that the context and the importance of this research can be better understood by the reader. The reason why only a broad overview is provided, is because of the comprehensive nature of both processes. A detailed and extensive discussion of various components of both processes will take too much time and space and might result in a loss of focus in this research. Therefore only the most relevant aspects were touched on and no claim of exhaustion of these processes is made by the researcher.

The chapter is divided into four sections. The first section will examine the role of evaluation in the teaching and education process, while the second section will analyse briefly the role of measurement as an essential component of the evaluation process. The third section will consider the implications of recent developments in the South African educational context on evaluation. Finally, the chapter will conclude with an examination of the implications recent developments in scientific thinking will have on the existing forms of evaluation and measurement.
2.2 EDUCATIONAL EVALUATION

2.2.1 The origin and historical development of educational evaluation

Because of the similarities between South African education and educational evaluation practices and those of the Western education system (Behr, 1988: 11, 12), no separate description of the development of educational evaluation in South Africa will be provided.

2.2.1.1 General development until 1950

The origin and history of the process of evaluation extend to the dawn of human existence. Popham (1993: 1) estimates the origin of the earliest forms of evaluation at about 4000 BC. Sax (1989: 6) traces the history of evaluation to 2700 BC where the Egyptians were already using measurement in three distinct dimensions. Most of the authors who have examined the historical development of educational evaluation (including Sax), lean strongly on the information provided by Dubois (1970) who highlights the use of measurement by the Chinese (Sax, 1989: 6; Cunningham, 1986: 3; Thorndike, Cunningham, Thorndike & Hagen, 1991: 2).

According to the information gathered from Dubois (Sax, 1989: 6), the Chinese had already developed formal oral tests of knowledge by 2200 BC. The Chinese emperor at the time instituted a process of competence testing for government officials which extended over a three-year period. The initial appointment of civil servants, as well as their continued employment, depended on their progress in this testing or examination process. Their knowledge of civil law, military affairs, agriculture, revenue, and geography was tested (Cunningham, 1986: 3). These Chinese testing practices served as a model for Western Europe and America because they later copied them in the 1800s (Thorndike et al, 1991: 3; Cunningham, 1986: 3).

According to some of the authors (Sax, 1989: 7; Hart, 1994: 2), the Bible also reveals examples of the use of tests and measurements. One example is that of Noah when
building the Ark (Sax, 1989:7). Noah had exact instructions in terms of the size of the
different parts thereof. Another example is provided by Hart (1994:2). She refers to
the event where Jephta of Giliad had to distinguish between his own men and the
defeated enemy troops. He used a pronunciation test in which the men had to
pronounce the word “Shibboleth” correctly.

According to Sax (1989:6) the first written tests were introduced into human society
as early as 202 BC. At that point in time the Han Dynasty used tests to measure civil
law, military-related subjects, agriculture, revenue, and geography. However, the first
written tests and examinations in schools, colleges and universities started in about
1500 AD and were first used by the Jesuits at St. Ignatius (Sax, 1989:8; Cunningham,
1986:3).

It was only in the early 1800s that educational measurement became a particular area
of study (Sax, 1989:8; Thorndike et al, 1991:3). The first printed tests made their
appearance in Boston in 1845 and consisted mainly of short answer essay-type
questions (Sax, 1989:8). It was also in 1845 that Horace Mann strongly advocated
the use of written tests and emphasized the need for test standardization in America
(Hart, 1994:2).

In 1900 the College Entrance Examination Board (CEEB) was formed in the United
States to standardize the secondary school curriculum and to make it acceptable to
colleges and universities. The CEEB administered their first examination in 1901,
consisting mainly of an essay format which included the following subjects: English,
German, Latin, Greek, History, Mathematics, Chemistry and Physics (Sax, 1989:8).

The early 20th century saw a growing need for the development of objective forms and
standards of measurement and evaluation as a result of the many inconsistencies that
were detected (Hart, 1994:3). This gave birth to the “scientific testing movement”.
This movement developed the standardized test in the form of a mass-produced
multiple-choice test which helped to minimize subjectivity. In the US the production
of standardized tests became an industry after World War I. By 1928 more than 1 000 such standardized tests were available there (Hart, 1994 : 3).

The 1930s and 1940s saw the introduction of the objective-based forms of evaluation spearheaded by Tyler in the US. Tyler’s influence will be described in 2.2.1.2.

The historical development of evaluation was particularly characterized by the influence of great scholars and the picture of the history of educational evaluation will be incomplete if their role is omitted. It is also important to note that educational and psychological measurement and evaluation developed side by side. Therefore in 2.2.1.2 a brief description of the contributions of some of the main role players in the history of educational and psychological evaluation will be provided. This list is however not exhaustive.

2.2.1.2 Important role players in the history of educational evaluation

- Wilhelm Wundt

He was from Leipzig in Germany and is regarded as the pioneer of turning psychology into a science. He placed great emphasis on measurement. The main forms of measurement used by him were mainly psychophysical (sensory, acuity, and reaction time) (Cunningham, 1986 : 4).

- Sir Francois Galton

The actual development of psychological testing began in the latter half of the 19th century under the guidance of Sir Francois Galton, a cousin of Charles Darwin of England (Cunningham, 1986 : 2; Sax, 1989 : 8). The absence of sophisticated measurement techniques made it difficult for him to assess the higher-level mental functioning of human beings. He was forced to use measurement instruments similar to the ones used by Wundt (Cunningham, 1986 : 6). Galton tried to establish norms and standard scores in testing. He can be regarded as the one who laid the foundation
for the development of the correlation coefficient. He shared this achievement with Karl Pearson (his biographer) and with Axel Oehrn (Cunningham, 1986: 6; Sax, 1989: 8; Thorndike et al, 1991: 3). Galton’s most important contribution is his emphasis on individual differences and the effect these have on people’s behaviour. This aspect forms a cornerstone of psychological measurement (Cunningham, 1986: 6).

- James McKeen Cattell

Cattell introduced Galton’s forms of testing to the US. He later also introduced his interest in psychological testing to the University of Pennsylvania and Columbia (Cunningham, 1986: 7).

- William Rice

In 1897 Rice, from Boston, used the first written spelling tests in public schools. Rice argued that too much time was spent in schools on spelling drills, and advocated that time should be used for teaching science. He demonstrated that the amount of time spent on spelling drills did not in fact correlate with high achievement in spelling. The importance of Rice’s study lies in the fact that it was one of the first instances where tests were used to evaluate the curriculum and to make a curriculum decision (Thorndike et al, 1991: 3).

- Edward Lee Thorndike

He was the author of the first book on mental and social measurements in 1904. Later, in 1910, he produced the first handwriting scale. He also, with the involvement of his learners, managed to develop educational achievement tests (Sax, 1989: 8).

- Alfred Binet

Binet, a child psychologist in Paris, became interested in the measurement of human abilities. He rejected the psychophysical measurement of Galton and Wundt and
promoted the measurement of higher-level mental processes. In 1905, with the cooperation of Theodore Simon, he developed the first intelligence test aimed at measurement higher-order mental functioning. He developed three measurement scales in 1905, 1908 and 1911. Binet’s main contributions to the measurement process were as follows:

- He focused attention on the assessment of higher-level mental functioning.

- He summarized the assessment of various aspects of cognitive functioning into a single score.

- He introduced the use of criterion groups (Cunningham, 1986 : 7).

• Louis Terman

Terman published the Stanford-Binet test in 1916. This test, a translation of Binet’s work, provided the first well-standardized and carefully developed test of intelligence which proved to be a more refined translation. Terman’s version combined test results with IQ scores in addition to mental age scores. The Stanford-Binet test was revised twice, in 1937 and 1960. However, after the appearance of the Wechsler tests, its use began to decline. The Wechsler tests, produced by David Wechsler, a chief psychologist at Bellevue Psychiatric Hospital, consisted of a series of subscales which enabled psychologists to make more detailed interpretations (Cunningham, 1986 : 7, 8, 269).

• Arthur Otis

He was a graduate learner of Terman and was instrumental in developing two tests on the request of the American Army for the purpose of inducting new soldiers during the First World War. The names of the two tests were the Army Alpha, which was used for soldiers who could read and the Army Beta, which was used for those who could not read. These two tests were significant in that they were ideally suitable for group
testing as opposed to the individual testing at which all existing tests at that stage were aimed. These tests had a great impact on ability testing because they led to the development of commercially published tests which included intelligence, achievement and personality tests (Cunningham, 1986:8).

- Ralph W. Tyler

Despite the long history of evaluation, the term “evaluation” only had its origin in the 20th century and its authorship is ascribed to Tyler (Norris, 1990:16). Prior to Tyler’s work, evaluation was regarded merely as measurement and testing.

As professor at Ohio State University in the US, he started in the 1930s with an eight-year study on the evaluation process. His study resulted in two conclusions, namely:

- that educational evaluation is not, in fact, the appraisal of learners, but rather of the quality of an educational programme;

- that evaluation should be objective- or goal-based, in other words, the evaluation process should aim to determine to what extent the goals and objectives of a particular educational programme were achieved.

Tyler’s findings influenced evaluation thinking for a long time (Popham, 1993:2).

- Benjamin Bloom

Michael & Michael (1995:2,3) mention the role of Benjamin Bloom who worked closely with Tyler. Bloom was concerned primarily with the development of a taxonomy of educational objectives. With the help of colleagues such as Engelhart, Furst, Hill, & Krathwohl, he managed to develop the well-known hierarchical structure of cognitive processes with knowledge at the lowest level and evaluation at the highest. The effects of this taxonomy on the development of the evaluation process are noteworthy. It helped define the process of clearly formulating instructional
objectives which in turn provided clear direction of what was to be evaluated in any evaluation process.

- Lee J Cronbach

In 1963 Cronbach wrote an article on the role of evaluation in curriculum development. His main argument was that educational evaluation should focus on the decisions faced by curriculum specialists in their efforts to develop the curriculum. Furthermore he stressed that evaluation activities should not focus on comparisons between programmes, but rather on the degree to which a particular programme promoted its desired outcomes (Popham, 1993: 3).

From the above description of the historic origin and development of educational evaluation as well as some of the key role players, the following conclusions can be drawn:

- First, educational evaluation is a very old process and probably had its origin in the dawn of human history.

- Second, evaluation developed gradually from a simple, optional practice into a highly sophisticated and compulsory field of study.

- Third, a number of individuals made major contributions to the development process of educational evaluation.

- Finally, because of the long history of educational evaluation and its growing compulsory nature in educational institutions, educational evaluation appears to be an essential component of the education and teaching process.

With these provisional conclusions in mind, the modern composition and nature of educational evaluation as well as the new trends that characterise the modern scene will next be considered.
2.2.2 New trends in educational evaluation

In the previous section a description of the historical development of educational evaluation until approximately the middle of the 20th century was provided. In this section the situation in the second half of the 20th century up to the current situation will be described.

The second half of the 20th century saw the appearance of various conceptual developments, under which the following resort: criterion-referenced testing; norm-referenced testing; item-response theory; the formulation of clear instructional objectives, etc. (Sax, 1989 : 9). The first two developments will be further discussed in 2.3.2.1 and 2.3.2.2.

The late 60s and early to middle 70s was characterized by an attitude of optimism towards the evaluation process by educators in the US. A more voluntary involvement, as opposed to the earlier state-imposed conditions in the early sixties, in the evaluation process by educators appeared evident (Popham, 1993 : 4). Policy makers in particular regarded the evaluation of educational programmes as most important. They saw such programme evaluation as a determining factor in deciding whether a particular programme should be abandoned or retained (Popham, 1993 : 5).

The latter half of the 20th century also saw the introduction and use of the computer not only in the development and administering of tests, but also in the marking (Sax, 1989 : 9). Merrill, Tolman, Christensen, Hammons, Vincent & Reynolds (1986 : 8, 9), for example describe the role of the computer in the education process as threefold, that is as tutor, tool and tutee. The first role especially dealt with the evaluation process, because as a tutor, the computer presents information to the learner; confronts the learner with questions and problems which must be answered; evaluates the learner’s response(s) according to specified criteria, and determines the further actions depending on the evaluation outcome of the learner’s response(s).
Apart from these positive developments in the evaluation process, the modern scene has also seen growing criticism against the current evaluation processes (Sax, 1989: 11; Hart, 1994: 6,7; Garcia & Pearson, 1994: 335 - 337; Cangelosi, 1990: 3; Frith & Macintosh, 1988: 79). Hart especially opposes the use of standardized tests. She identifies for example three points of criticism, namely:

- the presence of numerous flaws in those tests,
- the fact that those tests only measure the learners' test-taking ability,
- the fact that those tests are contrary to their own purposes because they corrupt the very process they are supposed to improve.

There are many other voices of criticism against the existing evaluation processes, but according to Sax (1989: 11) this criticism led to more and deeper research into the process of evaluation. It is probably also this criticism that led to a growing plea to move away from trivialized or rote learning to alternative and more process- or performance-oriented evaluation (Hart, 1994: 7; Van Koller, 1994: 51,52). The essential difference between the traditional or product-oriented forms of evaluation and the so-called alternative or authentic forms of evaluation that is advocated so strongly today, is that in the traditional evaluation methods, the emphasis was on rote learning, retention and repetition of learned information, while alternative evaluation focused on the skills mastered by learners in the learning process. The evaluation process therefore should involve learners in tasks that are worthwhile, significant and meaningful to learners (Van Koller, 1994: 51; Hart, 1994: 9).

This move away from the traditional forms of evaluation/measurement seems to be the most recent tendency in the field of evaluation. This feature corresponds with the most recent shifts in scientific thinking, especially with regards to post-modernism (Grebe, 1991: 128 - 132; Morrow, 1990: 47 - 55). This new direction in scientific thinking will be examined in 2.12.
This movement in evaluation circles also appears to be active in the South African education context (Van Koller, 1994: 52-54). The developments in South African education are however not only linked to common tendencies in global education, but also to changing political conditions in South Africa. The latter aspect in particular caused an urge for more accelerated change. In fact, the need for a complete transformation in the nature of the total education and evaluation processes in South African educational institutions in the light of the radical political and educational changes in South Africa, i.e. a new political order in South Africa; the new united education department under central control, etc., is increasingly being demanded by various parties. Samuels (1992: 3-7) for example emphasizes the need to address the numerous racial disparities and imbalances caused by the previous apartheid-based education system in South Africa. He sees it as the particular responsibility of tertiary educational institutions in SA to help eradicate those imbalances and disparities and to transform the educational process in South Africa.

It does however appear as if this call of Samuels has already been heard in tertiary education circles, because Pavlich & Orkin (1993: 10, 11) report that attempts to redress the imbalances and disparities of the past are already under way in some universities in South Africa. Their study has found that the following developments were detected in such universities:

- An emphasis on a move towards the development of certain skills so that learners will be able to meet the expectations of different faculties.

- A move towards becoming more reflexive and more critical towards their traditional teaching and learning strategies.

- A striving to create teaching and learning environments that respond favourably to the new learner composition at all universities, i.e. different race groups at previous racially exclusive institutions, learners from disadvantaged backgrounds, etc. This is seemingly implemented by means of academic support initiatives as well as by academic development programmes.
All the above-mentioned activities will have decided consequences on the evaluation methods used in those educational institutions, because it will no longer be fair and acceptable to use the old rote learning forms of evaluation which characterizes most educational institutions in South Africa (Grebe, 1991: 130, 131).

In conclusion, it would therefore appear that the dominant feature of the current educational evaluation scenario is a movement away from traditional and product-based forms of education and evaluation to the more process/performance-based forms of education and evaluation. This feature is also present in the South African educational context. However, the extent to which these developments have been implemented in South African institutions still remains to be seen.

With the above information in mind, the following is an attempt to describe the exact role of evaluation in the educational process.

2.2.3 The essential role of evaluation in the educational process

The role of evaluation/measurement seems to be indispensable in the education process. Evaluation can, in fact, be regarded as one of the three pillars on which the total education process rests, namely:

- Teaching

- Course/curriculum objectives

- Evaluation

This assertion is possible when one considers the fact that evaluation is regarded by a great number of authors as a prerequisite for the teaching process to succeed. Evaluation is also probably the only suitable manner in which to establish whether and to what extent the aims and objectives of an educational programme were achieved.
In light of these comments, an attempt will be made to illustrate the integrated position of evaluation in the educational process.

2.2.3.1 Evaluation as a prerequisite for the teaching process to succeed

Sax (1989 : 12) argues that teaching by its very nature requires continuous value judgements to be made by, for example, the teacher, the school administration, the parents and even the learners themselves. These value judgements can only be made by means of evaluation. Sax also stresses that it is essential that anyone thinking about education should recognize the interdependence of teaching and evaluation.

Connors (1983 : 16) sees the integration of evaluation with the teaching process as an essential requirement and that such evaluation must promote the learning process. He states: "So a good test teaches and good teaching involves testing".

Hopkins & Antes (1990 : 1) see the ascertainment of how well learners are learning as an integral part of the teaching-learning process and they claim that the evaluation of learner progress has been a part of the total education process since its inception.

Gronlund & Linn (1990 : 3,4) see evaluation as a systematic process that plays a significant role in effective teaching. According to them tests and other measurement instruments serve as a complement and supplement to the teacher's methods for obtaining information about learners. These measurement and evaluation procedures help provide a more comprehensive, systematic and objective base of evidence on which instructional decisions can be based.

Gronlund & Linn also provide a vivid illustration of how evaluation procedures can assist in the teaching process. According to them the following are examples of the major instructional decisions educators will normally encounter in their teaching
practice and some applicable evaluation/measurement procedures that will enable the teacher to make the correct decisions:

- “How realistic are my teaching plans for this particular group of learners? (Scholastic aptitude tests, past record of achievement.)

- How should the learners be grouped for more effective learning? (Scholastic aptitude and achievement tests, past record of achievement.)

- To what extent are the learners ready for the next learning experience? (Readiness tests, pre-tests over needed skills, past record of achievement.)

- To what extent are learners attaining the minimum essentials of the course? (Mastering tests, class projects, observation)

- To what extent are learners progressing beyond the minimum essentials? (Periodic quizzes, general achievement tests, observation)

- What types of learning difficulties are the learners encountering? (Diagnostic tests, observation, pupil conferences)

- Which learners are underachievers? (Scholastic aptitude tests, achievement tests)

- How effective was my teaching? (Achievement tests, learners’ ratings, supervisors’ ratings)” (Gronlund & Linn, 1990: 4,5).

These examples as well as the views of the other authors quoted in this section, provide a very clear indication of the integrated nature between teaching and evaluation and the essential role that evaluation must play in the education process as a whole.
2.2.3.2 Evaluation of the extent to which instructional objectives have been met

Gronlund & Linn (1990 : 8, 9) claim that the main purpose of classroom instruction, and any other instruction for that matter, is to help learners achieve a set of intended learning outcomes. Only by identifying and stating these intended learning outcomes or instructional objectives in a clear way beforehand, can the instructor provide clear direction to the teaching process. This is where the essential role of evaluation becomes very distinct, because evaluation is the essential tool whereby it can be determined whether those learning outcomes/objectives have in fact been achieved and to what extent. It is for this purpose that tests and other measurement instruments are designed.

Sax (1989: 13) confirms the views of Gronlund. He states that instructional objectives provide both educators and learners with goals to aim and strive for, while measurement and evaluation give the educator the opportunity to judge how well those objectives were attained. Sax further states that instructional goals in themselves imply the way in which such attainment of goals will be evaluated.

Fransman (1989 : 110) argues that the aspect of evaluating to what extent the instructional objectives have been realized is not only true in terms of short-term (formative) evaluation, but also in terms of long-term (summative) evaluation. He illustrates this by stating that an examination paper, which is mainly a summative evaluation/measurement instrument, reflects the total of lesson or course objectives.

When examining these different views, it would seem as if, without clearly stated goals and objectives, the determination of learners' academic progress and achievement is impossible. Without an evaluation of whether those goals have been attained or to what extent they were attained, the total teaching process will become meaningless and a waste of time.
In conclusion, it would appear as if evaluation can rightfully be regarded as the "third leg" of a trio consisting of teaching, setting of objectives and evaluation. With this clear description of the crucial role of evaluation in the education process as a basis for facilitating better understanding, some of the other purposes for which the evaluation process is used will be described in the next section.

2.2.4 Evaluation and the curriculum

It appears as if evaluation forms an integral part of curriculum planning and development. Just as evaluation forms a crucial component of the teaching process, as illustrated in the previous section, it plays a similar crucial role in the curriculum planning process.

Henson (1995 : 6) admits that there is little uniformity in terms of defining the concept of curriculum and provides a list of different definitions. However, he provides the following description of curriculum:

- "A curriculum usually contains a statement of aims and of specific objectives; it indicates some selection and organization of content, it either implies or manifests certain patterns of learning and teaching, whether because the objectives demand them or because the content organization requires them. Finally, it includes a program of evaluation of the outcomes" (Henson, 1995 : 4).

This central role of evaluation is also confirmed by Lawton (1989 : 14) who illustrates the curriculum planning process as follows:

- "Analyse the situation
- Define objectives
- Design the teaching-learning programme
- Interpret and implement the programme
- Assess and evaluate".
It appears therefore that the central role of evaluation in the curriculum, is to ensure that the aims, goals and objectives are achieved.

However, Lawton (1989: 13, 14) sees the role of evaluation also to be that of a process in which the “fragments of learning” are put together to ensure a full learning experience.

Posner (1995 : 230 - 232) distinguishes between different curriculum perspectives. Each perspective according to him, determines the main focus or emphasis of the evaluation component of every curriculum. Posner describes the following perspectives:

- A traditional perspective: Here the evaluation focus is mainly on reproduction of facts, basic skills mastery and the acquisition of societal values. The main aim of evaluation in this instance is to determine to what extent did the curriculum succeed in transmitting the desired facts, skills and values. Evaluation methods include standardized tests, classrooms recitation, assignments and reproducing what the teacher wants.

- The experiential perspective: The main focus in this instance is to ensure, by means of the learning experiences that a continuous process of learner development takes place. Evaluation methods are aimed at determining the long and short term effects of experiential programmes on learners. The focus is not only on outcomes, but also on the intrinsic value learners attach to these learning experiences. Evaluation methods are experience- and career-based.

- The behavioural perspective: In this instance, the main emphasis is on the performance of skills. Evaluation methods are concerned with determining whether the behaviours intended by the curriculum, were indeed acquired by learners. Evaluation methods mainly used or prescribed in this instance, are those that will manifest the behaviours acquired by learners. Examples provided by
Posner are observational checklists, practical examinations, but even paper and pencil tests.

- The perspective based on the structure of the disciplines: The main focus is obviously on the structure of a particular discipline. Evaluation methods used in this instance focus on the extent of knowledge acquired, the nature of learner inquiry and the conceptual structure of the subject contents. The types of evaluation methods used are very problem-, interpretation- or experiment-based.

- A cognitive perspective: The main emphasis seems to be on understanding and the fostering of thinking skills. Evaluation methods aim at monitoring to what extent learners have learned basic concepts in a meaningful way and acquired basic problem-solving skills. Evaluation methods normally used by this perspective are: clinical interviews, evaluation of learners' problem-solving skills, concept-mapping, and so forth (Posner, 1995 : 230 - 232).

Posner (1995 : 228) also points at the role of evaluation in the curriculum process as that of determining or indicating to what extent the goals and objectives of the curriculum were realised. He states that evaluation methods are mainly outcomes-based.

It does however appear as if there is a move away from the role of evaluation only at the end of the process, but continuously as an integrated part of the process. This is where the concept of continuous evaluation comes to the fore. Proponents of continuous evaluation argue that the evaluation process forms an integrated part of the curriculum as well as the teaching process and should therefore be reflected throughout the curriculum (Gearheart & Gearheart, 1990 : 82, 83; Gordon & Bonilla-Bowman, 1996: 36).

In chapter 5, section 5.6, the BM curriculum will be briefly analyzed in terms of its aims and objectives and the evaluation methods used in order to realise those aims and objectives.
2.2.5 Various other purposes of educational evaluation

The evaluation process is used for a number of other purposes in addition to the main purposes, illustrated in 2.2.3. A long list of purposes can be compiled which will be impossible to discuss in detail in this research. Therefore only a few examples of the varying uses of evaluation are listed to provide an idea of the multiplicity of uses of evaluation.

The following authors (Gronlund & Linn, 1990: 9; Sax, 1989: 13; Ebel & Frisbie, 1986: 17; Popham, 1993: 5; Hopkins & Antes, 1990: 30, 31; Van Koller, 1994: 7, 8; etc.) mention the following different purposes of educational evaluation:

- Evaluation is the instrument used to determine the extent to which the objectives of an educational programme were achieved.

- It can be used as an instrument to improve learners' learning and progress because of the feedback it provides.

- Evaluation can be used to determine the entry level of learners at new educational institutions they wish to attend.

- It is used to establish the usefulness of an educational programme.

- It helps indicate the presence of gaps or shortcomings in a programme or method with the intention to rectify the situation.

- It provides educational decision-makers with essential information on educational phenomena which will impact on the decisions they must make.

- It gathers information about a wide range of learner characteristics as a basis for making decisions.
• It helps to accumulate records of progress.

• It provides information from which educators can obtain insights into their own effectiveness.

• It helps to inform other educators who must make decisions about their learners.

• It makes it possible to compare the progress of learners under different educators.

• It can be used to compare new teaching materials with old ones.

• It can be used as an incentive to learning and an aid to remembering.

• It can be used to inform parents about the progress of their children.

• It informs employers or institutions of higher education about learner attainment.

• It can be used to diagnose learners' strengths and weaknesses.

• It can be used to assess the extent to which learners have benefited from a course of instruction.

• It evaluates the effectiveness of methods of teaching.

• It predicts learners' future performance.

• It places learners in the most beneficial educational situation.

Despite the variety of the purposes or uses of the evaluation process, there appears to be unanimity among various authors about one thing and that is that the aims of any
type of evaluation should correlate with the overall goals and aims of that particular institution or training programme (Norris, 1990: 17; Rowntree, 1990: 305; Gronlund & Linn, 1990: 8, 9; Sax, 1989: 13; Ebel & Frisbie, 1986: 17; Fransman, 1989: 110).

However, this wide and diverse range of purposes for educational evaluation forms a basis for the existence of different models of educational evaluation. Wolf (1990: 61) states that a fundamental view of the basic purpose(s) of educational evaluation held by supporters of each evaluation model, forms the basis of the differences between the various models. Because of the important role such educational evaluation models play in the total evaluation scenario, a brief description of some of the major educational evaluation models will be considered in the next section.

2.2.6 Educational evaluation models

The term “model” in educational evaluation context can have two meanings, namely:

- First, a more prescriptive type of model, which is the most common type. This type of model includes a set of rules, prescriptions, prohibitions, and guiding frameworks. These aspects are used to prescribe a good or proper evaluation and how it should be performed.

- Second is a more descriptive type of model. This type of model normally describes, predicts or explains evaluation activities and provides a range of possibilities for conducting evaluation (Alkin & Ellet, 1990: 15).

Alkin & Ellet (1990: 15) see as the basic areas of focus of most models, the following:

- First, the evaluation of the effectiveness of the teaching and learning processes and the curriculum.

- Second, studying the activities, practices and policies designed to facilitate the effectiveness of the various educational processes.
Below is a short summary of the most common educational evaluation models:

- Naturalistic evaluation

Naturalistic evaluation refers to evaluation processes in which information is acquired in a systematic manner and by means of naturalistic oriented methods. One of the best examples of naturalistic research methods is anthropological or sociological fieldwork. This form of inquiry focuses on extended personal observation by researchers of a particular study group or culture in their naturally occurring circumstances. The central principles of the naturalistic orientation are:

- a primary focus on the views and perspectives of the participants, both during the research process as well as the time of reporting on it;

- an approach based on holism which makes a concerted effort to indicate and understand the context of particular events within the larger background of interdependent relationships within a particular group;

- an approach wherein the evaluator becomes the primary instrument for the collecting of data and its analysis (Dorre-Bremme, 1990: 66,67).

- Illuminative evaluation

This model of evaluation aims at investigating and interpreting a variety of educational phenomena. These include an investigation into the experiences of participants in the evaluation process, different institutional procedures as well as specific management problems, in such a way that the parties for whom the evaluation is intended, will be able to recognise the usefulness of such evaluation for them. The following are some of the basic principles of this model:

- A primary concern with phenomena of organized complexity;
- An assumption that events are normally interrelated;

- A belief in the probability of multiple causality;

- A belief that it is their task to unravel the complicated webs of causes and effects of particular events.

Illuminative evaluators normally portray the outcomes of their evaluation processes in a map-like fashion in which they attempt to illustrate the various relationships between the various components of a particular system (Parlett, 1990: 68, 69).

- Transactional evaluation

Transactional evaluation examines the question of how people who are involved in new programmes deal with discrepant personal needs and aspirations as well as institutional demands and sanctions. Traditional organizational theory testifies that an attractional dimension exists between the nomothetic (institutional) and the ideographic (personal). This model proposes the combination of adversary groups in the same evaluation team. The model which is therefore based upon the transaction between such two adversary teams, differs from its predecessors in four respects, namely:

- intents

- methods

- roles

- composition

The intent of this model is aimed at the expansion of the data collection process by including social and interpersonal obstacles to change. It uses mainly conventional
methods and a certain type of questionnaire of which the items are obtained both from evaluation specialists as well as their critics. The evaluator's role is adapted to such an extent that it could make room for substantial interventions as well as intimacy (Rippey, 1990: 73).

- **Intrinsic evaluation**

This is a type of evaluation model that focuses on the evaluation of curriculum materials and proposals and is more concerned with the intrinsic nature of such curriculum material than the effects thereof. This term was introduced by Scriven in 1967. Scriven believed that pure empiricism was impossible and that the emphasis should rather be on the evaluation of the goals of a programme. According to him the intrinsic component of an evaluation study should also include two types of analysis, namely:

- a consistency analysis in which the differences between the various types of goals can be made known;

- a content analysis with the purpose to evaluate the content of a particular curriculum (Eraut, 1990: 74).

- **Responsive evaluation**

The fundamental principle of this type of model is its intention to be responsive to various issues during an evaluation process. Such issues refer in particular to the feelings and perceptions of the participants present at the evaluation site. Responsive evaluation focus on aspects such as:

- the various activities of a programme

- its uniqueness
It requires a delay in the whole process in order to allow the evaluators to become more familiar with the programme and its context. During these delays the goals of the evaluation can also be adapted to become more relevant (Stake, 1990: 75).

- Adversary evaluation

As the title states, this type of evaluation model makes use of two adversary forms of data collection, namely quantitative and qualitative data gathering. This form of evaluation can best meet needs such as providing comprehensive information on the most prominent features of particular social situations as well as information about the interrelatedness of such features. The adversary approach was introduced by people such as Owens (1973) and Wolf (1975). The main consequence of such joint use of two opposite forms of data gathering is that two independent evaluations will in fact take place, and that, in turn, will make possible more valid judgements and decisions (Clyne, 1990: 78).

- Judicial Evaluation

The central purpose of the authors of this model was not to replicate legal procedures in the evaluation process, but rather to ensure more discreet evaluation practices by the use of the law, where applicable. The model was initiated in the early 1970s as a method which would use procedures from both court proceedings and administrative hearings in the field of law. The purpose of this was to promote greater appreciation among educational decision-makers for the complexity of their programmes. It was also intended to help such educational decision-makers to base their decisions on more complete and valid information (Wolf, 1990: 79).
• Delphi Technique

This technique was named after the oracle at Delphi in ancient Greece. It is a communications process which enables groups to solve problems by reaching consensus without in fact being in face-to-face contact with one another. Data is posted to the members of the evaluation team and they have to provide comments on that data, but on an anonymous basis (Uhl, 1990 : 81).

Popham (1993 : 22 - 49) arranged the various models, including those mentioned above, under the following categories:

- Goal attainment models
- Judgemental models emphasizing inputs
- Judgemental models emphasizing outputs
- Decision-facilitation models
- Naturalistic models

Choppin (1990 : 8) and Alkin & Ellet (1990 : 16) want to ascribe exclusive use of the term “evaluation” to programme evaluation, while the evaluation of learner progress is termed “assessment”. According to Alkin & Ellet (1990 : 16) most evaluation models tend to focus on programme evaluation and not on the evaluation of learners’ progress or on educators’ effectiveness. They use the term “assessment” to denote learner and teacher evaluation and ascribe the term evaluation only for the evaluation of educational programmes. Choppin argues that the term “evaluation” should only be used to denote abstract entities such as programmes, curricula and organizational variables (1990 : 8). This view is strongly contested in this research for the reasons stated below.
First, when the definition of evaluation as provided in chapter 1, section 1.8.1, is analysed, it appears that the process of the evaluation of learner progress or achievement complies with all the elements of the inclusive definition of evaluation, namely:

- **Measurement**: In the evaluation of learner achievement, tests and examinations are used to measure the extent to which they have mastered or attained certain course objectives or skills.

- **Assessment**: By analysing and interpreting the results of the measurement process and by awarding a particular mark the learner's achievement is assessed.

- **Judgement**: After assessing learners' achievement, a judgement is made in terms of:
  - whether a learner has passed a course;
  - whether the learner has passed the course well, average or poorly.

The aspect of judgement particularly in terms of evaluation features strongly in the definitions of evaluation of a number of authors (Cangelosi, 1990: 25; Lawton & Gordon, 1993: 82; Barrow & Milburn, 1990: 124). Salvia & Ysseldyke (1995: 36) make the following statement about the evaluation of learner achievement/progress:

"...for example, report-card grades represent the teacher's judgement of a learner's academic progress during the marking period."

Second, when the various definitions of the different authors in chapter 1, section 1.8.1, are examined, it is clear that not one of the definitions specifically links
evaluation to programme evaluation. In fact, the definitions allow for the inclusion of a wide scope of aspects that can be evaluated and therefore their suggestion that it refers exclusively to programme evaluation becomes a contradiction of the very definition of evaluation.

- Third, the question arises that if the term "assessment" is used to indicate the evaluation of learner achievement, does assessment contain the element of judgement which seems to be a strong feature of the evaluation of learner achievement as indicated above? An analysis of what the process of assessment entails, as in 1.8.1, indicate that assessment is primarily concerned with the analysis and interpretation of data in order to establish its worth. There is no mention of the aspect of issuing judgements in the process of assessment. Salvia & Ysseldyke themselves (1995 : 5) whose work is titled "Assessment" defines assessment as "the process of collecting data for the purposes of making decisions about learners". However they do not see assessment as the actual process of making the decisions, but rather as the process undertaken to make decisions possible.

- Finally, a number of authors use the term "evaluation of student achievement/progress" as if it is a generally accepted custom (Salvia & Ysseldyke, 1995 : 16 - 18; Sax, 1989 : 533 - 562; Gronlund & Linn, 1990 : 3 - 6; Ebel & Frisbie, 1986 : 13). The fact that this term appears to be in common use further strengthens the argument that the use of the term "evaluation" for learner achievement/progress is fully justified.

In view of all these considerations, the contention in this research is that the term "evaluation" can be used to denote the evaluation of learner achievement/progress and that it is not scientifically justifiable to use the term exclusively to denote programme evaluation.

At this stage it will be useful to distinguish between the main forms of evaluation. This will be done in the next section.
2.2.7 The main forms of educational evaluation

A distinction is drawn between two basic forms of evaluation, that is formative and summative evaluation (Thorpe, 1988: 9,10; Rowntree, 1990: 302, 303; Bergman, 1981: 54; Agar, 1988: 223; Cangelosi, 1990: 2,3; Lewy, 1990, 26 - 28; Ebel & Frisbie, 1986: 13, 14; Van Koller, 1994: 10, 11). A brief summary of what the two forms of evaluation entails according to these authors will be discussed in this section.

- Formative evaluation refers to the process of evaluating the progress made towards achieving programme goals, during its implementation. The aim is not to measure the effectiveness of a programme, but to identify any changes it might require in order to improve it and make it more likely to achieve its goals in the long term. Formative evaluation therefore seems to refer to the continuous process of evaluation throughout the year or study period. Probably the most common instrument for the purposes of formative evaluation in educational institutions is the test.

- Summative evaluation, on the other hand, is concerned with the evaluation of the final outcome of a programme or period of study. It is normally conducted on completion of a programme or course of study. In educational institutions this normally takes the form of year-end examinations which are normally concerned with the rounding off of a programme.

Although these two forms of evaluation are “entrenched” in the traditional evaluation contexts, it would appear that there is an increasing shift in emphasis away from summative evaluation towards formative evaluation (Van Koller, 1994: 55). This shift in emphasis occurs by letting the marks obtained in the formative evaluation process count equivalent value and even more than the mark obtained in the summative evaluation process in terms of learners’ final marks. This shift seems to be completely warranted in the light of the alleged shortcomings in the summative evaluation instruments.
With this brief description of the two basic forms of evaluation, the first part of this chapter is completed. In the second part of the chapter, the role and function of the measurement process will be described.

2.3 EDUCATIONAL MEASUREMENT

2.3.1 Measurement: The essential component of evaluation

Most of the authors, who discuss the topic of measurement, point to the crucial role that measurement fulfils in the process of evaluation.

- Hopkins & Antes (1990: 4) regard measurement as a prerequisite for evaluation because, according to them, sound evaluation judgements are based upon systematically collected information. Measurement is thus the process that provides that information so essential for making the evaluation process possible, because without collected information to analyse and interpret, no judgements are possible and therefore the evaluation process cannot take place (Hopkins & Antes, 1990: 29).

- Fransman (1989: 110, 111) also underscores this view. He states that the evaluation process is based on measurement and that it is impossible to achieve the objectives of evaluation if there is no sound quantitative basis available.

- Thorndike et al (1991: 8) and Ebel & Frisbie (1986: 14) also emphasize the essential role of measurement in the evaluation process. They see this role as the provision of the information on which the total evaluation process and the process of decision-making, which is a consequence thereof, depends.

In the light of the views of these authors, it would appear to be an undeniable fact that measurement is a crucial and essential component of the evaluation process. Logic makes it clear that no realistic and acceptable judgement, another essential part of evaluation, can take place if there is no information available on which to base that
judgement on. With such clarity on the matter, a description of the basic forms of educational measurement be discussed next.

2.3.2 Basic forms of educational measurement

It appears as if tests (formative evaluation) and examinations (summative evaluation) are the most common forms or instruments of measurement in most educational institutions.

A test is described by Ebel & Frisbie (1986: 16) as a set of questions, each with its own correct answer. These questions must be answered by each individual examinee either orally or in writing. They further state that tests are a subset of the total of measurement tools, which can include both quantitative and qualitative forms of measurement. Another important aspect they mention, and which might be an issue of common understanding, is that tests should contribute towards summative evaluation. In other words, tests as an instrument for formative evaluation, should be employed for the purposes of improving the process of summative evaluation.

Because of the great number and various types of tests and the fact that they represent formative evaluation, only the most prominent types of tests will be described briefly. Examinations, as basic form of summative measurement, will be discussed extensively in chapter 3 and will therefore receive no further attention in this chapter.

Some of the most prominent categories of tests are:

- Norm-referenced testing
- Criterion referenced testing

Most of the various types of tests (of which a list will be provided) will fall somewhere in or between these two categories. Therefore only the categories will be discussed while only a list of the various types of tests will be provided.
2.3.2.1 Norm-referenced testing

Sax (1989: 9) describes norm-referenced testing as an attempt to measure individual differences. He claims the essential characteristic of this type of test is the intention of comparing one learner's progress with that of another. The focus according to him is on the differences between individual learners in as far as achievement, intelligence, interests, attitudes or personality are concerned.

Thorndike et al (1991: 195) see the purpose of norm-referenced testing as making inferences about how much a learner has learnt in comparison with others.

Hopkins & Antes (1990: 21) place the emphasis on the comparison of tests performances of different learners with that of others in the same class or standard. They also mention the use of the results of other learners who wrote the test previously as external criteria for the evaluation of the performance levels of other learners.

Mehrens & Lehmann (1991: 16) describe norm-referenced tests as the process where the scores of one individual learner are compared to those of other individuals who are then seen as the norm-group for that particular learner.

In essence then, it would appear as if norm-referenced testing is based on the comparison of learners' scores with those of other learners which forms the norm-group for every learner involved.

2.3.2.2 Criterion-referenced testing

According to Sax (1989: 9) criterion-referenced tests are aimed at the measurement of a specific domain of knowledge. In contrast to norm-referenced tests, this type of test focuses on the individual's mastery or attainment of specific subject matter and objectives. Criterion-referenced tests are not concerned about the comparison of learners with one another, but with the achievements of individual learners in relation to stated syllabus objectives.
Mehrens & Lehmann (1991: 16) see criterion-referenced tests as a comparison of every individual learner's performance against a specified behavioural domain or criterion of proficiency. The focus in this case is on what a learner can in fact do, in other words the skills the learner has mastered.

Thorndike, et al (1991: 195) state that criterion-referenced tests enable educators to make absolute decisions about whether a learner has learned specific course contents. This is in contrast to more relative decisions that results from norm-referenced tests.

Hopkins & Antes also underscore this explanation of criterion referenced tests (1990: 15). They emphasize that the educator should have established precisely what knowledge and behaviour define the particular domain, and what test activities will portray the achievement of that particular domain.

2.3.2.3 Other test types

The following authors provide the following list of test types (Sax, 1989: 19 - 23; Ebel & Frisbie, 1986: 16 - 30, 126 - 196, 267; Thorndike et al, 1991: 160, 192; Gronlund & Linn, 1990: 13, 291 - 315):

- Teacher-made tests: These tests are constructed locally by educators themselves and normally administered only to the classes that are taught by that particular teacher.

- Standardized tests: These tests are constructed by test specialists working with curriculum experts and educators and are normally administered on a national basis.

- Individual tests: These are tests designed to be administered to one person at a time.
• Group tests: Tests that are administered to a number of people at the same time.

• Objective tests: Tests that have clear and unambiguous scoring criteria and independent scorers can agree on the number of marks that should be awarded to every answer, i.e. multiple-choice, true or false, matching items.

• Subjective tests: Tests that have a vague scoring criteria and independent scorers can easily disagree on the number of marks to be awarded.

• Power tests: Tests that have generous time limits and allow learners to attempt every item in the test.

• Speed tests: Tests that have severe time limits but which are normally very easy and few learners normally make errors.

• Verbal tests: Tests that emphasize the use of language as the basic means of responding to test items or questions.

• Non-verbal tests: Tests that de-emphasize the role of reading in testing.

• Performance tests: Tests that require examinees to perform a task other than merely answering questions.

• Non-language tests: Tests that are administered in the form of gestures and require no language facility on the part of the examinee.

• Sample tests: Tests that are used to measure only a partial aspect or sample of the learner's total behaviour.

• Sign tests: Tests that are used diagnostically to distinguish one group of individuals from another.
• Achievement tests: Tests that measure to what extent the learners have achieved the cognitive mastery of the knowledge in which they were instructed.

• Essay tests: Tests of which the answers will be provided in the form of an essay.

• Diagnostic tests: Tests of a very specialized type used to help detect certain recurring learning difficulties in learners in order to find a remedy.

• Aptitude tests: Tests that measure a learner's ability to learn new tasks.

With this description of some of the basic forms of measurement, the researcher proceeds to an examination of the essential requirements that any measurement instrument must comply with in order to be accepted.

2.3.3 The essential requirements for a measurement instrument

There are a number of requirements that any measurement instrument must comply with. Examples of such requirements are:

• Fairness

• No ethnic or gender bias

• Elimination of subjectivity in marking or scoring

• Accurate interpretations of test results

• Generalizability

• Applicability

• Suitability
These are merely a few examples and many other requirements can be added. The remainder of this section will be devoted to examining some of the most essential requirements that any measuring instrument must comply with, namely validity, reliability, accountability and legitimacy.

2.3.3.1 Validity

The term “validity” insofar as the process of measurement is concerned, is commonly understood to mean that a certain measurement instrument is able to measure what it is supposed to or intended to measure. For example, if the purpose is to measure learners’ spelling accuracy in a certain language, the test/examination used for that purpose must measure only the learners’ spelling accuracy and nothing else (Thorndike, 1991: 91; Hopkins & Antes, 1990: 5,6; Salvia & Ysseldyke, 1995: 162 - 170; Sax, 1989: 295 - 301; Van Koller, 1994: 25).

Salvia & Ysseldyke (1995: 162) focus the attention on the fact that validity is not in actual fact a property of the test itself, but rather a property of test-based inferences. They also distinguish between three main forms of validity, namely:

- Content validity, which means the extent to which test items actually represent the domain which is to be measured.

- Criterion-related validity, which indicates the extent to which a person’s performance on a criterion test can be estimated from that same person’s performance in a previous criterion test.

- Construct validity, which indicates the extent to which a test measures a specific trait or characteristic (Salvia & Ysseldyke, 1995: 163 - 170).
Other authors such as Thorndike (1991: 124-140), Gronlund & Linn (1990: 51) and Sax (1989: 294-301) also discuss these three types of validity. Gronlund & Linn (1990: 47) especially emphasize the need for appropriateness of the inferences or interpretations made from test scores and other measurement results.

2.3.3.2 Reliability

The term "reliability" in the context of educational measurement appears to denote the extent to which a particular measurement instrument is capable of rendering similar results on a continual basis (Salvia & Ysseldyke, 1995: 136; Thorndike et al, 1991: 91; Hopkins & Antes, 1990: 5; Gronlund & Linn, 1990: 47).

Salvia & Ysseldyke (1995:136) see the primary purpose of reliability to be the possibility of generalizing. According to them, there are three areas of generalization and these lead to three forms of reliability, namely:

- Reliability for generalizing to different scorers called interrater or interscorer reliability.
- Reliability for generalizing to different times called stability or test-retest reliability.
- Reliability for generalizing to other test items called alternate forms or internal-consistency reliability.


- The test-retest method in which the same test is used twice in succession with the same group of learners with a given time interval between the two administrations. The results of these two testing events are correlated and the correlation coefficient provides a measure of stability.
• **Equivalent forms method:** This method uses two different but equal forms of the test (parallel or alternate forms). The two forms of the test are written by the same group of learners in close succession and the results obtained are correlated. This correlation coefficient provides a measure of equivalence. In other words these tests indicate the extent to which both forms of the test are measuring the same aspects as intended.

• **The split-half method:** This is when a single test is administered to learners and afterwards a split is drawn between the marks of the even numbers of the test and the odd numbers of the test. Each learner will then receive two scores and these two scores are then correlated with each other and provide a measure of internal consistency. The coefficient indicates to what extent consistent results were obtained from the two halves of the test.

• **The Kuder-Richardson method:** This method also provides a measure of internal consistency, but does not require a splitting of the test results. One of the formulas, the Kuder-Richardson 20, is based on the proportion of persons passing each item and the standard deviation of the total scores. This method indicates to what extent the items in the test measure similar characteristics.

In conclusion, some authors regard validity as the most crucial requirement of a measurement instrument (Thorndike, et al, 1991 : 123; Cunningham, 1986 : 100; Hopkins & Antes, 1990 : 6; Ebel & Frisbie, 1986 :71). They see reliability as a necessary, but not sufficient condition for validity because tests can be reliable without being valid. In other words, a test might be highly reliable in that it continuously generates similar results, but the test results might not bear any resemblance to what they were intended to measure. The main purpose therefore is to establish whether a test is indeed measuring exactly that for which it was intended.

The next section will describe some other essential requirements to measurement instruments that are increasingly being emphasized recently.
2.3.3.3 Accountability/Legitimacy


With regard to accountability the argument is that educational institutions are accountable to learners, parents, potential employers and the public concerning the education and training of learners or potential employees. For that reason, educational institutions should use measurement instruments that provide clear records of the academic progress and achievement of learners. The evaluation methods/measurement instruments should also provide learners with regular feedback in order to allow for remediation where necessary. However, there is a contra-argument that contents that educational institutions are not the only ones who should be held accountable. Learners should increasingly take greater responsibility for their own learning and become less dependent on the educational institution. No matter what view is supported, the fact is that educational institutions should ensure that the evaluation methods/measurement instruments they use, will comply with the requirements of accountability.

The issue of legitimacy is mainly concerned with the aspect of adequate involvement of all stakeholders in the teaching/learning scenario. With regard to evaluation/measurement this issue implies that all stakeholders be involved in the development and implementation of evaluation methods/measurement instruments. Legitimacy, for example, demands active learner involvement in the learning/evaluation processes and decisions. Whenever new or alternative methods are introduced, it should enjoy the support of the majority of stakeholders. See section 4.4.6 for a more elaborate discussion of these aspects.
With this description of the essential requirements of measurement instruments, the next section will be devoted to a description of the exact nature of educational measurement.

2.3.4 The exact nature of educational measurement

2.3.4.1 The process of measurement

Thorndike, et al (1991 : 9) state that measurement in any field involves three common steps:

- The identification and definition of the quality of the attribute to be measured;

- A determination of the set of operations by which the attribute can be isolated and displayed;

- The establishment of a set of procedures for the translation of observations into quantitative statements of degree or amount.

Thorndike, et al (1991 : 13) see quantification of attributes as the third step in the measurement process. According to them the process of measurement involves the assigning of numbers to objects or people and those numbers represent how much of the attribute is present in the person or thing. The first step in this quantification process is to establish a set of rules on which the assignment of numbers will be based. Those rules will then make it possible to answer questions such as: How many, or how much. They further state that there are a number of advantages attached to the use of numbers, of which two are:

- Quantification enhances the preciseness and efficiency of the communication process.
The application of mathematics to observations helps to make those observations more meaningful.

Hopkins & Antes (1990: 10) underscore this explanation of the measurement process by Thorndike, et al. They see the primary purpose of assigning numbers to attributes of objects, persons or events to be in order to point out differences in amount. They make the following examples of the question “how much” normally asked in the measurement process:

- How much spelling achievement does a particular person have?

- How much history achievement does another learner have? etc.

According to them, quantification provides answers to such questions.

Cunningham (1986: 14) however, points out one shortcoming of the quantification process in educational measurement. The problem according to him, lies in the fact that when a single number is assigned to a construct, an assumption is made that a particular trait with a unitary meaning exists. This is not however the case, because a number of factors are involved. He illustrates this by means of a reading achievement test. Such a test represents the sampling of a range of behaviours and not only a single one. According to Cunningham, in most cases, the person using the scores from such tests are not even aware of all the factors that are in reality being assessed. Cunningham’s argument raises the question of whether any test or examination in education for that matter can really be regarded as valid, because although a test might measure a particular trait as intended, so many other traits are simultaneously also being measured which were not intended by the author of the test.

Another shortcoming in the process of quantification is the problem of measurement error. The concept of measurement error refers to the impossibility of rendering measurement of any nature that is completely accurate and without any error. Cunningham (1986: 12) for example describes measurement error as an inevitable
factor in the measurement process. Thorndike et al (1991: 92) state that measurement error can sometimes be positive and in such a case it will make a certain score too high, while if it is negative, it will make the score too low. This is one aspect that educational measurement has in common with physical measurement, according to Hopkins & Antes (1990: 9). They state that no measurement procedure can produce a measure that is completely exact. The best is to come close to exact. However, according to them, the doubt associated with the uncertainties of measurement of an educational construct is much greater than in the case of physical measurement.

The existence of the aspect of measurement error brings to the fore another term in educational measurement, namely the aspect of true scores. With this term is meant the real actual score that an item measured would have obtained if measurement error were completely absent. Thorndike et al (1991: 92) for example explain this by means of an illustration of the weighing process. They state that the score of a person being weighed will consist of the true score, that is the person's real and constant weight, plus some error of measurement. Cunningham (1986: 12) on the other hand states that the term "true score" can have more than one meaning in the context of educational measurement. It can refer first to what is called "classic true score" which is a term used in certain formal psychometric theories. Second, it can refer to a hypothetical perfectly accurate score free of any measurement error and which stands opposed to the obtained score of a person or object which normally includes measurement error.

Unfortunately the nature and purpose of this research does not allow space to examine the abovementioned issues in the educational measurement process in more detail. The second aspect to be addressed in this section is the question of what exactly is measured in educational measurement.

2.3.4.2 What is measured in educational measurement?

Hopkins & Antes (1990: 7) state that the attributes most commonly measured in classrooms are cognitive achievement and cognitive aspects of skill development.
They recommend that the measurements of achievement be limited to achievement in well-defined domains of knowledge or skills. They also make it clear that it is not the human being that is being measured, since that is not really possible, but rather certain properties of a human being, for example, achievements in certain areas; height; weight; motivation; creativity, etc.

Thorndike et al (1991 : 9) and Cunningham (1986 : 13, 14) also agree on this point. Thorndike et al make it abundantly clear when they say that it is not a person or thing that is being measured, but a quality or attribute of a person or thing. They provide additional examples of attributes measured in educational evaluation, namely the intelligence of a school child, as well as the achievement of a learner in a particular subject, etc.

Hopkins & Antes (1990 : 10) further emphasize the importance of reaching consensus between educators, and in the mind of an individual educator, about what exactly is to be measured. The need for a common understanding in this regard is very crucial in order to clearly, or as far as possible, define the attribute and to interpret correctly the findings produced in the measurement process.

The aspect of attaining common agreement is however not so easily achieved. Thorndike et al (1991 : 10 - 16) for example, mention that it in many occasions is not possible to obtain complete agreement insofar as the definition of physical attributes are concerned. In the case of education and psychology that works mainly with abstract and non-physical attributes, the problem is even more complex. Thorndike et al provide some examples of the types of dilemmas educators can be faced with regarding the definition of a particular construct. In the case of human intelligence for example, questions such as the following may arise:

- What is meant by the term "intelligence"?

- What types of behaviour can be characterized as intelligent?
• Should the construct be defined in terms of ideas and abstract concepts or in terms of dealings with concrete objects?

• Will the construct refer mainly to behaviour in novel situations, or to behaviour in familiar and habitual surroundings?

Thorndike et al further state that although groups of people may have a general idea about something, for instance intelligence, there are still points on which they will disagree as each individual tries to make a precise definition. They give the example of educators who are confronted daily by the problem of assessing learner performance. An understanding of the term “performance” however, is closely linked to the way in which it is defined and assessed. Only to the extent that educators agree on the precise definition of the term “performance” will their assessments have any comparable meaning. If one educator, for example regards as good performance the effective recall of facts, while another understands the concept to mean effective application of principles in a subject, they will to some extent evaluate different aspects under the same title (Thorndike et al, 1991: 12, 13). It appears as if this is an inescapable problem with regards to the definition of almost all psychological and educational constructs.

It is not only the question of correct definitions on which common agreement is essential, but also the relevance of a construct (Thorndike et al, 1991: 10). Thorndike et al stress the importance of determining the relevance of a construct to be measured, because in the absence of such relevance, such measurement will be useless and a waste of time. The traits to be measured should therefore be relevant to the decisions that will be based on them and an educator dare not measure traits merely because they are easy to assess.

2.3.4.3 How are attributes defined in educational measurement?

Another aspect that needs to be examined further is the way in which attributes or constructs are defined in educational measurement. In educational measurement
frequent use is made of what is called operational definitions of attributes (Hopkins & Antes, 1990: 11, 12; Thorndike et al, 1991: 12).

However, before the concept of operational definition can be defined, another term needs some clearer definition, and that is the term “construct”. This term is used in educational measurement to denote more abstract and difficult to observe properties of people, for example intelligence or personality (Thorndike et al, 1991: 9). Since most of the aspects that are measured in education are more of an abstract than physical nature, this term is commonly used.

The term “operational definition” refers to the process in which, on the one hand, the definitions of any concept determines what will be accepted as reasonable and relevant operations thereof, while on the other hand, the operations devised in the measurement instrument to elicit or display the attribute, constitute in a practical sense the definition of that attribute (Thorndike et al, 1991: 12).

An example of this process of operational definition is provided by Hopkins & Antes (1990: 11, 12). They cite the example of a “paper and pencil-test” used to measure a learner’s history achievement. According to them, such a test will define history achievement by means of the tasks (operations) that will be demanded from the test-taker, while the tasks or operations will in fact constitute the definition “history achievement”.

Thorndike & Antes (1990: 12) state that an operational definition gives meaning to a construct because the tasks assigned to learners will specify what attribute is being measured. This aspect is demonstrated in the following way: They state that popularity is determined by the number of choices each member of a group of learners receives from their peer group when the following question is posed to them: “Who would you like to have as partner for the science project?” When interpreting this example, it would appear as if the term “popularity” is the construct, while an attribute of popularity would be preference by others.
2.3.4.4 Difficulties in educational measurement as compared to physical measurement

Educational evaluators are faced with a number of problems when it comes to the process of measurement. Thorndike et al (1991: 14, 15) summarize those problems into three categories:

- Problems with regard to the selection and definition of attributes;

- Problems relating to procedures to elicit the relevant attributes of constructs;

- Problems in terms of the establishment of the equality of units in educational measurement.

Hopkins & Antes (1990: 7 - 11) argue that it is easier to measure physical attributes than cognitive and affective ones and that measures for physical attributes are easier to interpret. They identify the following problems:

- The fact that there are no commonly accepted interpretations for measures on mental ability, scores of achievement tests, ratings of attitudes, level of motivation, etc.

- In the case of physical measurement, equal units of measure exist, while in the case of educational measurement this is not the case. For example, in physical measurement any centimetre equals any other centimetre, while in the case of educational measurement devices used to measure attributes rarely have units that will allow for equal measures between different measurement opportunities (even of the same aspects).

- Educational measurement instruments are based on assumptions. They state that the use of tests to measure non-physical (abstract) aspects, relies on assumptions. Such assumptions involve the view that all tasks, for example, are equal to others for the test-taker. A concerted effort is also normally made to create conditions
that will meet the assumption. They give the example of a spelling test in which the educator will try to assure that the range of difficulty between the words is relatively small in comparison to the range of difficulty in the dictionary. Hopkins & Antes further state that although the quantification of the test (counting how many words were spelled correctly) allows a definition of amount, it still does not establish the equivalence of tasks as is assumed.

• There are two factors that form the main sources of doubt insofar as measurement in psychology and education is concerned:

- The attributes being studied are very difficult to define because of the difficulty in conceptualizing them. Definitions and measurement of educational attributes require in the first place that constructs should be formed in order to try and explain those abstract concepts. This makes it very difficult to reach agreement on what attributes actually entail or which attributes must be measured.

- The instruments used to measure educational constructs are not yet as developed and sophisticated as the ones used in physical measurement.

Thorndike et al (1991 : 14 ) capture this same point and illustrate it by means of the following example: They pose the question of how it will be possible for an educational evaluator to demonstrate or prove that one arithmetic problem is equal in the amount of arithmetic ability to another arithmetic problem. They also mention that the educational evaluator will be forced first to try and define a particular construct in order to be able to provide units and quantification. According to them, although educational evaluators can, for example, call a certain task successfully completed, or a word that was correctly defined, or an arithmetic problem that was solved, etc., equal to any other task in the series and count the total number of successes for an individual, such an evaluator will still not be able to provide adequate evidence of the equivalence of different test tasks.
From the arguments of the various authors above, it becomes very clear that the measurement process in education is a complex and even problematic one. It would appear as if one of the central problems is that of developing measurement units that will have more or less equal values. The possibility of locating a solution in that regard seems very remote in the foreseeable future. A possible alternative solution to this problem might be to develop measurement instruments that will clearly elicit the particular attributes educational evaluators intend to measure and in so doing, helping to lessen the dependency on exact measurement units. One of the objectives of this research is that its results and findings should contribute towards a reduction of or even solution to this problem.

These last comments conclude this section. The next section will be devoted to a reflection on the possible implications of recent developments in the South African educational context.

2.4 IMPLICATIONS OF RECENT DEVELOPMENTS IN THE SOUTH AFRICAN EDUCATIONAL CONTEXT

A number of important developments in the South African education context recently will probably have a very great influence on the evaluation practices used in educational institutions. These developments are:

- First, the creation of a National Qualifications Framework (NQF);

- Second, the introduction of an Outcomes based Education approach in South African schools;

- Third, the technology-enhanced learning initiative.
2.4.1 The NQF and SAQA

2.4.1.1 What is the NQF and SAQA?

The National Qualifications Framework is a body that is in the process of being established. The main purpose for this body, was to create a national framework in South Africa for all education and training processes and qualifications to establish national, uniform standards for any qualifications to be awarded (SAQA, 1997: 2, 3; Office of the President, 1995: 1, 2).

The desire to establish a body such as the NQF was kindled by widespread discontent among people with the standards and quality of both the education and training processes in South Africa. Quality and standards are therefore some of the main points of emphasis in the NQF. The issue of standards and quality is also emphasized strongly in the 1997 Draft White Paper on Higher Education, (Ministry of Education, 1997: 6, 11).

SAQA on the other hand is the authority established by the South African Qualifications Authority Act, 1995 for the purpose of overseeing the development and implementation of the NQF as well as the furtherance of the objectives of the NQF (SAQA, 1997: 4, 5; Office of the President, 1995: 3).

2.4.1.2 The influence of the NQF on evaluation practices

The fact that the NQF is aimed at the establishment of uniform standards and a maintaining and improvement of quality in the service and products to be delivered by educational and training institutions, will affect all facets of the higher education / assessment processes. The evaluation / assessment methods used in higher education should ensure that that any learner receiving a particular qualification complies with the requirements and criteria of that qualifications.
An aspect such as evaluation/assessment criteria forms an important element of the composition of a unit standard, the term suggested to indicate the uniform standards that will be used across the board in education as well as training (SAQA, 1997; 9). Evaluation/assessment is also listed as one of the registration criteria for a qualification (SAQA, 1997: 16). In addition, it is stated that evaluation should form an integral part of the learning process and should consist of both formative and summative evaluation methods. It is also important to note that SAQA recommends the use of a variety of evaluation methods which includes written, oral, practical and simulated forms of evaluation (SAQA, 1997: 16).

With this brief discussion on the influence of the NQF on evaluation practices in educational institutions in South Africa, it is evident that evaluation methods used in future will have to comply with the requirements set by the NQF.

Another process that developed from the creation of the NQF and that will have a definite influence on educational evaluation practices, is the Education, Training and Development Practices Project which will be discussed in the next section.

2.4.1.3 The Education, Training and Development Practices Project

As one of the consequences of the development of a National Qualifications Framework, and the development of an integrated qualifications framework as well as the aim towards establishing uniform standards, was the launch of the Education, Training and Development Practices Project (ETD-practices project) by the National Training Board (National Training Board, 1997: 11). The overall aim of the ETD-practices project is to develop an integrated education, training and development field.

The other more specific aims of the ETD Practices-project are summarized as follows:

- The promotion of portability and mobility for learners between the different areas of ETD;
- The improvement of the quality of learning, both formal and informal learning;
• The development of a high level of competence among all ETD practitioners;
• The creation of learning and progression opportunities and avenues for ETD practitioners;
• The development of a standard-setting model that can assist ETD practitioners in the process of curriculum development. (National Training Board, 1997: 12).

The latest available report of the ETD-practices project that the researcher could obtain showed that the process has reached its third phase.

• The first phase of the project entailed a research process aimed at obtaining information such as: different groups of ETD practitioners; job opportunities or areas for practitioners; the different institutions that normally employ ETD practitioners; the perspective of practitioners on the issue of professionalism; and the views of practitioners on the suspect of standards within ETD.

• The second phase was concerned with the development of two standard-setting models.

• The third phase will involve a piloting process in which the two models would be piloted at various sites to determine their feasibility (McLean, 1997: 15, 16).

From this brief description on the main aims and processes of the ETD-practices project it is clear that this project is mainly concerned with the establishment of uniform standards on the field of Education, Training and Development. The implications of this project will most probably affect all aspects in the ETD-field, in particular the aspect of evaluation methods used in the education, training and development processes. Any evaluation/measurement instruments that do not comply with the requirements of the final standards, will no longer be acceptable.

In order to give a clearer perspective on the aspect of standards in technikon education, the body currently responsible for promoting and maintaining standards at technikons will be described in the next section.
2.4.1.4 SERTEC

The body currently responsible for the attaining and development of standards at technikons is the Certification Council for Technikon Education or SERTEC (SERTEC, 1996: 33). The SERTEC manual makes it very clear that it is SERTEC's duty to protect the education and examination standards in technikons. The aim of such protection is to ensure that there is compliance with prescribed standards of quality (SERTEC, 1996: 33, 34).

SERTEC was established in 1986 in terms of the Certification Council for Technikon Education Act, 1986 (Act 88 of 1986). SERTEC regards one of its main functions as ensuring that all technikons in SA comply with certain minimum standards (SERTEC, 1996: 1). The process in which compliance with these minimum standards is evaluated is by committees appointed by SERTEC visiting all academic departments at technikons at regular intervals (1996:1).

2.4.2 Outcomes-based Education (OBE)

2.4.2.1 What is OBE?

The introduction of outcomes-based education in SA is also based on the NQF. The term used by SAQA is "Critical Cross-field Education and Training Outcomes" or the shortened form, "Critical Outcomes" (SAQA, 1997: 6). SAQA, in the process of developing the NQF and based on comprehensive research, decided on this outcomes-based approach. The intention was to change all forms of education from an input focus to an outcome focus.

Outcomes-based-Education is defined by the Department of Education as:

- "a learner-centred, results-oriented design, based on the belief that all individuals can learn" (Department of Education, 1997A: 17).
Outcomes are defined as:

- "Outcomes refer to the specification of what learners are able to do at the end of a learning experience" (Department of Education, 1997B: 12).

Key elements of outcomes-based education, are the following:

- The focus is on the purpose and outcomes or results of the learning process (the what).
- The goals of education are stated very clearly.
- The contents of what is to be learned are stated very clearly.
- A variety of teaching, learning and evaluation strategies are used in order to accommodate all learners.
- Enough time and support are provided to all learners in order to realise the potential of all learners. (Department of Education, 1997A: 17).

The new requirements placed on evaluation practices in educational institutions, as a result of the OBE approach, is clearly stated by the Department of Education, namely:

Evaluation methods should entail the following characteristics:

- There should be clear links between the evaluation methods and the outcomes produced by it.
- The evaluation methods should form an integrated part of the teaching and learning process.
- A multiplicity of methods should be available which can ensure a comprehensive but balanced level of evaluation.
- Evaluation methods should be able to prove its validity.
- The evaluation methods should give all learners a fair chance, irrespective of race, religion or language.
- The evaluation methods should involve learners in meaningful activities.
• The methods should produce enough evidence to make it possible for educators to issue judgements based on that evidence.
• Evaluation methods should be adaptable to the unique nature of different learners and reflect their individual achievements.
• The evaluation methods should however also be able to be applied for a complete education institution.
• Evaluation methods should also aim to promote parent involvement in the learning processes of their children.

In conclusion, the information conveyed by the evaluation methods should be meaningful and useful to both learner and educator (Department of Education, 1997A: 30 - 32). The Department of Higher Education also recommends some newer methods for evaluation, for example: portfolio evaluation, performance evaluation, peer and self-evaluation, observational sheets, journals and so forth (1997A: 33 - 35).

A brief description of the Department of Education’s implementation plan for OBE will be discussed in the next section.

2.4.2.2 Curriculum 2005

The Department of Education’s plans for the new learning frameworks and programmes for primary and secondary schools and its proposed implementation dates, were announced in the Government Gazette No. 17724, on 19 January 1997 (Department of Education, 1997B: 1). It was clearly stated that this arrangement was made to in order to comply with the OBE introduced by SAQA for the purposes of the NQF.

Some of the main critical outcomes as intended by SAQA in Curriculum 2005 are the following:

• The identification and solving of problems by means of critical and creative thinking;
• The development of group work skills that will enable learners to function as a member of a team;

• Effective management skills;

• The collection, analysis, organisation and critical evaluation of new information;

• Effective communication skills by means of different modes;

• The effective use of science and technology and a responsible attitude to the environment;

• The development of a proper understanding of the related nature of various systems in the world (Department of Education, 1997B: 10).

Following this brief description of the essence of Curriculum 2005 the researcher will next consider a final development in the SA education scenario that might have an influence on future evaluation practices in education.

2.4.3 Technology-Enhanced Learning Initiative

The Department of Education recently published a discussion document on the aspect concerning the use of technology in education and training. The main aim with this document is apparently to develop a five-year strategic plan for the implementation of the Department's policy on technology in education (Department of Education, 1997C:1). The document is based on research done by the Technology-Enhanced Learning Investigation team who started their research in 1995.

The term "technology-enhanced learning" has apparently been formulated by the Department of Education and seems to refer to the use of any type of technology for educational purposes in any teaching/learning situation (Department of Education,
The main areas for the use of technology in education as distinguished by the Department, are the following:

- Technologies to assist in the provision and delivery of course material to learners.

- Technologies to be used supplementary to other teaching and learning processes.

- Technologies that can be used for the purpose of management and administration.

It seems as if the Department of Education is also encouraging the increased use of technology by all educational institutions, in particular distance education institutions. Some of the advantages of the use of technology are identified by the Department. These advantages are:

- It can help to improve the quality of education.

- It can help to redress the shortcomings and imbalances of the previous educational system in South Africa.

- It can help to develop new strategies for teaching and learning (Department of Education, 1997C:9).

It seems as if Technikon SA has anticipated this greater emphasis on the use of technology in the education process. Not only did TSA become connected to Internet and the World-wide Web, the global computer network (Boyle, 1997:17), but it has also started the development of a "Virtual Campus". The virtual campus is a system whereby TSA learners can, among other things:

- Receive courseware via TSA intranet;

- Receive on-line guidance by lecturers;
• Receive their tutorial letters and assignments via TSA intranet;

• Become involved in electronic discussion groups;

• Obtain all relevant information or contact with lecturers and tutors via TSA intranet (Van der Merwe & Lazenby, 1997: 1-8).

(The TSA intranet is an internal web station that only allows access to TSA staff and learners.)

The intention is also to, at a later stage, offer on-line measurement/evaluation for learner academic progress and achievement. This will include eventually formative and summative measurement/evaluation.

With this information as background, it seems as if technology-based evaluation methods will most probably also become one of the requirements for evaluation practices at TSA in future.

2.5 IMPLICATIONS OF RECENT SCIENTIFIC DEVELOPMENTS ON THE EVALUATION AND MEASUREMENT PROCESSES

Scientific thinking in the 80s and 90s is characterized by the “post-'s”; for example, post-structuralism, post-marxism and post-modernism. Modernism itself seems to be a “post-” itself, namely post-Enlightenment.

Cahoone (1996: 6) states that some major shifts in philosophical thinking occurred in the latter half of the 20th century. These shifts entail a break away from all traditional and established paradigms including modernism and positivism. Positivism or logical empiricism that dominated the first half of the 20th century was one of the established paradigms that received increasing criticism from modern scholars. The strong emphasis on the human logic as basis for all systemized knowledge was no longer acceptable to many scholars and philosophers. The new buzz -word in philosophical thinking appears to be the concept of post-modernism. Post-modernist thinking is
apparently a growing phenomenon in the current world context and its influence is already noticeable. This influence of post-modernism will have definite implications on various spheres in the global context, including the spheres of education and evaluation. As a result of the increasing influence of this new paradigm it is necessary to describe it briefly and to consider the possible implications it might have on educational evaluation processes.

2.5.1 What is post-modernism?

Post-modernism is described in various ways by various authors and it seems as if there is no uniform understanding or agreement on what the term entails. Cahoone (1996: 1) states that the term means to one group of philosophers a break away from all former paradigms, as mentioned above. For another group it entails an attempt aimed at the destruction of Western civilization by a group of left-wingers. The following definitions and views on post-modernism give a good picture of the differing opinions on what it actually entails.

Grebe (1991: 128 - 131) describes post-modernism in the following way:

- Post-modernism is a radical form of criticism against modernism and especially against the modern Western world.

- The concept of post-modernism, is difficult to define or to structure, because the movement itself rejects such definition or structuring.

- It represents a movement that levels radical criticism on a broad basis against different intellectual discourses. It proposes a new paradigm in which reality should be re-interpreted.

- Criticism of post-modernism include the contention that the "safe havens" of emancipation, that has reason as its basis, created by autonomous and emancipated man, has been turned upside down and has turned into an enslavement and
exploitation of man by its bureaucracy and technocracy. As a result of these developments a situation was created in which the legitimacy of the modern technocratic society as well as education, the instrument through which the aims and ideals of this society is to be realised, is questioned from different angles and contexts.

- Post-modernism has opposed specifically the following three fundamental principles of modernism:
  
  - The principle of autonomous reasoning (rationality)
  
  - The anthropocentric understanding of the subject of reason
  
  - The continuity model of historic progress

According to post-modern thinking, these fundamental principles are no longer valid.

Post-modernism appears to have two distinct legs: One is completely anti-modernistic and proposes a complete break with modernism, while the other leg proposes merely a broadening of various aspects of modernism, and not complete rejection.

The first leg of post-modernism is completely rejected by most modernist oriented scholars (Grebe, 1991: 131; Van der Walt, 1988: 192, 193) They regard it as an irrational and relativistic attempt, which cannot justify its own claims.

Van der Walt (1988: 189 - 191) describes post-modernism in the following way:

- The term "post-modernism" has its origin in areas such as architecture and literature and was later transferred to the field of philosophy and eventually to education.
Post-modernism is characterized by a scepticism and unbelief against certain modernist views and beliefs.

Post-modernism is a mental attitude characterised by a type of desperateness and a fundamental scepticism against the development of theories.

In post-modernism, anything can be added to anything without any historical context or ties. It refuses morale and politics.

In terms of pedagogics, it expresses a fundamental doubt about the claims of pedagogics being a science. It refuses the "pedagogization" of everyday life. It demands a return to everyday life, the real life which has been forgotten by the science of pedagogics.

Post-modernism is an attempt to strip off the masks, the facades and to reject the superficial.

It announces the fall of the principle of reason of modernity.

Post-modernist thinkers want to strip off all masks and distanciate themselves from the influence domain of the intellectuals.

They are nihilistic because everything in life has lost its meaning for its supporters. For post-modernists, there is nothing left in life for which to live. There is no more meaning to life and it is not even worthwhile to search for meaning.

However, man can still try to live life to the fullest by means of thoughts of the "here and now"; by identifying with the feelings of the moment and the needs of the self.

Post-modernists prefer to trust their own experiences of life and not a rationalistic legalisation of reality.
• Post-modernism is a rebellion against modernism and specifically against the "meta-theoretical legitimizing of knowledge" which in principle finds it difficult to accept that knowledge has its origin in experiences.

• Post-modernism turns its back on the dualistic theory of modernism (rationalism and the practical).

Morrow (1990: 50 - 55) in turn describes post-modernism in the following way:

• The crucial difference between modernism and post-modernism is that modernism regards particular representations as problematic, while post-modernism on the other hand "problematizes" reality. In other words, where modernism questions modes of representation, post-modernism in fact questions reality itself.

• Post-modernism is more of a mass concept, mass-media, television, videos, etc., while modernism is more of an elitist phenomenon.

• Post-modernism "decentres" the identity and culture of the working class.

• Post-modernism has already raised its head in the 1960s, an era characterized by a global political reawakening because in the sixties certain notions of modernism were already being challenged.

• Morrow also links post-modernism to the information revolution and to things such as television, videos, computers, etc., and sees this revolution even in less developed countries, such as Mexico and other Latin American countries.

• Morrow also states that the dilemmas of post-modernist culture cannot be avoided by any society.
These summaries of the descriptions of post-modernism by the various authors provide a basic understanding of post-modernist thinking. As initially stated, it would appear that post-modernism is, in essence, a reaction against modernism which, on the one hand, advocates a total rejection of modernism and, on the other hand, a broadening in the scope and principles of modernism. Despite the fact that some authors and thinkers reject post-modernism outright, it is clear that post-modernism is a force to be taken into account and that its influence is very noticeable in educational circles.

With these comments in mind, it is very important to examine the implications of post-modernism on the educational process as well as the evaluation process. This will be done in the next two sections.

2.5.2 What are the implications of post-modernism on education?

Barros-Mani (1991: 19 - 23) provides a vivid picture of the implications of the post-modernist paradigm on education in particular. Here are some of those implications he identified:

- The fact that the status of knowledge will change and is already in the process of changing. This will entail that anything that is currently regarded as knowledge will be abandoned if it cannot be translated into “quantities of information”, and that access to stored information as well as information networks will become the determining factor for future research.

- Barros-Mani also states that the old belief that learning and the acquisition of knowledge cannot be separated from the training and development process of the human mind, is an outdated concept. Knowledge is produced with the purpose of selling it, and is no longer an end in itself.

- Knowledge and learning are no longer primarily formatively inclined and will no longer hamper individual self-reflection. Instead, it has become more of a merchandise to be used for other purposes.
• Knowledge, because of its informative nature, has become a major factor in the global power struggle, so much so that it is possible in future that the fight between nations will be about information and knowledge and not about the access to raw materials and cheap labour as previously.

• The finding of truth is no longer the major aim of knowledge, but rather finding and establishing the best performance or input-output equation. This search for optimal performance forms a major thrust in the movement to change universities, especially in developing countries such as Latin America.

• In higher education there is already a shift away from truth for its own sake to performativity. This can be seen in certain higher education policies. The emphasis currently is on the reproduction and development of skills for the purpose of power and profit.

• Post-modernist developments might make it possible in future for the teacher's position to be taken over by computers of learning and information databanks.

• The education system in Latin America should move away from the system proposed by modernism towards the post-modernist model. The education system must become more mobile and flexible in order to adapt to the changing world and its growing needs.

• Learners should no longer be inculcated with knowledge contents, but rather with the language and skills to enable them to have access to the modern information systems and to be able to solve the current problems facing them.

• To implement such changes as proposed, fundamental structural changes and new policies will be required in universities.
• Educational institutions, because of the influence of post-modernism, are becoming more functional since the emphasis is no longer on the ends, but on the means or the skills required to accomplish those ends.

• Universities are faced with new challenges of which one of the main ones is the quest to educate learners to be able to face the demands of the new and unpredictable circumstances at this stage and in future. All these challenges will require a fundamental renewal in terms of teaching and research strategies as well as academic structures.

• Barros-Mani proposes the following three strategies for higher education to address adequately the challenges it faces:
  - Greater interdisciplinary education and research
  - A multi-disciplinary curriculum; and
  - A critical inclination

• Finally, he also proposes greater communication in tertiary education.

Grebe (1991: 129 - 131) also identifies a number of implications that post-modernism has posed for the education process:

• She mentions that certain post-modernist scholars, such as Mollenhauer, Giesecke, Lyotard, Rang, etc., have already started advocating an end to the existing modernist-based pedagogic movement.

• Some of the important questions that can be levelled against the modernist education system according to her, are:
whether educators can continue to teach and evaluate knowledge contents that might be irrelevant in a very short time to come;

- whether, in South Africa in particular, with its multicultural population composition, educators can still continue to work according to curricula based on Western modernist rationalism;

- whether education aiming at the perpetuation of modernism in the South African context with its unique problems, is still justifiable.

- What is required from educators, in the context of post-modernism, is first to reconstruct the shortcomings of the concept of rationality, the backbone of educational theory and practice, and second, to take a stance of positive criticism to determine, in a realistic way, which of the core elements of the enlightenment tradition, are still essential for education.

- One of the major shifts in education is the epistemological move away from the concept of closed knowledge systems to open and dynamic knowledge systems.

- Education should move away from the rationality or positivist model which believes that the growth of science is based on compiling more and more facts and that there are autonomous knowledge systems which must be mastered by the human being, to the post-modernist tendency of the management of knowledge and information.

- No longer should the emphasis in education be on the instruction of contents, but rather on the acquiring and mastery of essential cognitive and other skills that will enable learners to have access to relevant information and to be able to manage that information.
• Our education system, which includes the curriculum, teaching and evaluation, should move away from a primarily Western-based one towards a more pluralistic one which will be suitable for the pluriform situation in South Africa.

Van der Walt (1988: 191-193) seems to be strongly modernistic-oriented because he takes a strong stand against post-modernism and even calls the movement an irrational approach to the education reality. He argues that the criticism of post-modernism against the modernist pedagogization of society and the education reality is unconvincing. This is how he describes the implications of post-modernism on modern education:

• He sees as one of the consequences of the post-modernist movement, the fact that empiricism has once again taken root in the education process. The real world has become important again and learning through experience has become more important than pure academic understanding and concepts.

• Post-modernism wants education and teaching to return to its common, everyday nature which includes the ordinary problems, stereotypes and routines of the education process.

• Post-modernism therefore advocates a move away from the rationalist theory construction of education as a science, to a more reality conscious practice. In other words, a greater concentration on the "here and now" in the education reality.

• Van der Walt sees this as being irrational to a certain extent, because of the post-modernist view of not grounding education on a scientific basis, but on focusing on the education reality without being norm-bound.

• Another possible consequence of such post-modernist influence, is the fact that pedagogics will no longer feel itself bound to a fixed system of educational concepts.
All the abovementioned comments reflect very clearly the profound implications that post-modernist thinking is already having on the education process as a whole. In summary it would appear as if the following are some of the central implications of post-modernist thinking on education:

- A move away from the belief of closed systems of knowledge to more open systems of knowledge and open access to them.

- A move away from teaching and learning knowledge quantities towards a greater focus on performance and skills.

- A changing role of knowledge becoming more a merchandise to be used for purposes other than only being an end in itself.

This description of the implications of post-modernist thinking on education creates a better context for an examination of what the implications of post-modernism will be on the educational evaluation process. The next section will attempt to clarify these implications.

2.5.3 What are the implications of post-modernism on the evaluation process?

As a result of the emphasis on greater experience-based education by post-modernism (Van der Waldt, 1988 : 191), evaluation methods and instruments will have to be turned towards the evaluation of more than merely theoretical content-based knowledge. In fact, both Grebe (1991 : 131) and Barros-Mani (1991 : 20) advises a move away from the evaluation of learners' memory abilities to the evaluation of practical skills and performance.

Grebe (1991 : 131) further states with regards to the South African context, that the evaluation process should in accordance with the teaching process itself cater for the pluriform or multicultural nature of its society. According to her, it is clear that the
current, modernist-based evaluation methods do not comply with this requirement of the South African society.

From an examination of the statements, views and descriptions of post-modernism by the various authors, the following list of implications on the educational evaluation process seems to be logical:

- There must be a move away from the existing, primarily content-based, forms of evaluation and measurement to more suitable forms of measurement and evaluation.

- Evaluation methods should be more skills-based and therefore make room for the evaluation of actual performance so that the skills employed can be portrayed.

- Evaluation forms must be developed that will cater adequately for the nature of pluriform societies such as those in South Africa.

- More subject-specific forms of evaluation should be produced to provide the best possible recording of the achievement of learners in each particular discipline.

- The examination of alternative forms of evaluation should receive urgent attention in all educational institutions world-wide and in South Africa in particular.

Many other implications can probably be added to this list, but for the purposes of this research the above-mentioned implications represents some of the major ones insofar as education and educational evaluation is concerned. This list of undeniable implications of the post-modernist era in which society and education in particular, finds itself, provides an even clearer understanding of the motivating causes that prompted this research.

These closing comments bring to an end this chapter. The final section is a brief summary of its contents.
2.6 SUMMARY

This chapter started with a brief overview of the historical development of the evaluation process and some prominent new tendencies in educational evaluation. Thereafter the crucial role of evaluation in the total teaching and education process was examined and then the main purposes of evaluation was determined. A brief description of some major evaluation models and the most basic forms of evaluation were then given. The second half of this chapter was devoted to examining the role and function of measurement in the evaluation process. The crucial role of measurement, as an essential component of evaluation, was illustrated. Thereafter, the basic forms of measurement, the essential requirements measurement must comply with and the exact nature of educational measurement were analysed briefly. Following that, certain recent developments in the South African educational scene were briefly examined. Finally, a brief description was given of the possible implications of current scientific thinking on the education process in general and on evaluation in particular.

The next chapter will be concerned with an in-depth scrutiny of the existing summative measurement instruments used at Technikon SA and elsewhere. This chapter will include a description of these instruments and a literature-based examination of their advantages and disadvantages.
CHAPTER 3

CONVENTIONAL SUMMATIVE MEASUREMENT INSTRUMENTS

3.1 INTRODUCTION

In chapter 2 the researcher gave a broad overview of the processes of evaluation and measurement. A description of the origin and historic developments of evaluation and some modern trends were discussed. The essential role of evaluation in the teaching process as well as the main forms and models of evaluation were illustrated. The phenomenon "measurement" was also described. The basic forms and requirements of measurement were discussed. The scope of educational measurement and its limitations were also described. The main implications of recent developments in the South African educational context on evaluation/measurement processes and practices were identified. The chapter ended with a brief exposition of the influences and effects of current scientific thinking on educational evaluation and measurement processes.

In this chapter the researcher will attempt to describe the main forms of summative measurement instruments used on a common basis world-wide. Then he will sketch a broad picture of the most common or popularly used measurement instruments globally. A brief description of the various measurement instruments used for Business Management at tertiary institutions in South Africa and at TSA will be given. Finally, a scrutiny of the most common measurement instruments will be undertaken in order to investigate the authenticity of the claims of numerous shortcomings leveled against conventional summative measurement instruments in chapter 1, section 1.3.

The first aspect that needs to be clarified in this chapter is the concept of conventional summative measurement instruments.
3.2 WHAT IS MEANT BY CONVENTIONAL SUMMATIVE MEASUREMENT INSTRUMENTS?

First, the concept of conventional summative measurement instruments in this research refers to those measurement instruments used in educational evaluation for the greatest part of educational history.

Second, conventional also implies those measurement instruments that were/are in common use in various educational institutions. In other words, the use of such instruments is not the exception, but rather the rule and they are used in a number of institutions at primary, secondary and tertiary level. These also include some newer forms of measurement initiated in the 1970s and 1980s which have become globally acceptable. The emphasis is on common usage.

Third, the term "conventional measurement instruments" is also used as the opposite of alternative measurement instruments (see chapter 1, section 1.8.5).

The terms "summative" and "measurement instrument" were already defined in chapter 1 (section. 1.8.2 and 1.8.4) and the same meanings apply in this instance.

3.3 A BASIC DESCRIPTION OF THE MOST COMMON CONVENTIONAL MEASUREMENT INSTRUMENTS

In order to help the reader better understand why the measurement instruments examined in this chapter are regarded as conventional, a brief exposition of the vast extent of social measurement instruments is supplied.

3.3.1 Different methods/techniques of measurement

A wide variety of measurement techniques exist. However, only a brief illustration of the main categories will be given, after which a detailed analysis of the most common category will be provided.
Frith & Macintosh (1988: 52) and Cox (1994: 84) present a basic summary of the various measurement categories. The main categories identified by these authors are as follows:

- Written examinations
- Practical evaluation
- Oral evaluation
- Aural evaluation
- Questionnaires
- Coursework
- Laboratory reports
- Observation

Bergman (1981: 183-240) discusses the following additional categories:

- Grading
- Rating
- Social measurement

These measurement instruments mentioned above are regarded as categories in this research because most of them can be divided into smaller variations which can
function as independent measurement instruments. A brief illustration of these various categories with their variations or examples will be supplied:

- Tests (described in chapter 2, sections 2.3.2.1 - 2.3.2.3)

- Examinations
  - Standard conventional examinations
  - Essay-type questions
  - Structured essay-type examinations
  - Case studies (written)
  - Open-book examinations
  - Pre-set questions

- Practical evaluation
  - JMB Physical Science O-level
  - SREB Home Management
  - SREB Art and Craft
  - SREB Biology
  - In-basket tests
  - Role-playing account, and so forth

- Oral evaluation or Vivas
  - Speeches
  - Debates
  - Discussions
  - Dialogues
  - Short talks, and so forth
  - Other subjects

- Aural evaluation
  - Foreign/different languages
  - To evaluate music
- General issues, and so forth

- Questionnaires
  - Learner attitudes to the subject
  - Learner attitudes to the method of instruction
  - Learners' study-methods, and so forth

- Coursework
  - Projects
  - Field work, and so forth

- Observation
  - Direct observation
  - Indirect observation, e.g. video-recordings

- Grading
  - Letter grades
  - Pass-fail grades
  - Checklists, and so forth

- Rating
  - Ability judgement scales
  - Performance scales
  - Personality scales
  - Agreement scales
  - Similarity scales
  - Frequency of observation scale
  - Normative rating scale
  - Value judgement rating scale
  - Trait rating scales, and so forth
- Social measurement
  - The personal interview
  - Case histories
  - Interest surveys
  - Autobiographies
  - Sociograms, and so forth


This list of categories is by no means exhaustive and each of the variations justifies a full description. However, a description of all the abovementioned categories and variations falls outside the scope of this research. This exposition was merely presented to give an indication of the broad and extensive scope of measurement instruments or techniques used in social measurement. It is also important to note that not all these categories can be used for summative evaluation nor are all of them normally used for educational measurement purposes.

As a result of the numerous variations of measurement categories and the focus of this research, it was only possible to concentrate on the category(-ies) in most common use in education globally as well as in Business Management departments at tertiary institutions in South Africa. The most dominant category appears to be written forms of measurement and therefore a brief description of the common forms and variations is in order.

The most dominant form of written evaluation appears to be the traditional and very familiar written examination process. The majority of the authors consulted (Frith & Macintosh, 1988:54-81; Bergman,1981:183-240; Cawood et al., 1988:120-136; White, 1983:132) place special emphasis on examination as a very popular measurement instrument/technique for summative evaluation purposes. As a result of the apparent prominence and common use of examinations in education in general and
in Business Management at TSA and other tertiary institutions in South Africa in particular, as established in this research (see sections 3.4, 5.3 and 5.4), the researcher decided to focus on the written examination only.

For the purposes of this research only those variations of the written examination used for summative evaluation were analysed. The variations were:

- The standard conventional examination
- The essay-type examination
- The structured essay-type examination
- The written case study
- The open-book examination
- The pre-set question or seen examination

Although the extent of popular use of these examination forms differ, they are regarded by a number of authors as some of the most commonly used forms of examination (Frith & Macintosh, 1988:54-81; White, 1983:132; Bergman, 1981:8 82, 83, 88-124, 183-240; Cawood et al., 1988:120-136; Duminy & Söhnge, 1987:117, 125 - 135; Rowntree, 1990: 306,307 - 313; Mathews, 1985 : 104; Habeshaw et al, 1993 : 35-42, 65 - 67). For the purposes of this research, these six variations of the written examination (which are regarded as six measurement instruments) will be taken for further scrutiny and analysis. Such a scrutiny will start of with a description of each measurement instrument.

A description of the various types of examinations as well as the different types of questions, is very useful because it will help the reader to have a better understanding of the global context of examinations as well as the scrutiny of the most common ones.
a little later in sections 3.4 and 3.5. The first measurement instrument to be described is the standard conventional examination.

3.3.2 The Standard Conventional Examination

The concept, examination, can briefly be defined as a process of testing the knowledge or ability of, for example, learners by means of oral or written questions (Sykes, 1982: 335). The emphasis in this instance of course is on the written examination.

The concept, standard conventional examination, is a concept specifically used in this research which is intended to mean the more general and traditional form of examination in which there is:

- a variety of questions (both lower and higher cognitive in nature; different types, such as objective/subjective and so forth);
- no prior knowledge about the exact wording or composition of the questions by learners;
- no textbook or any study material allowed inside the examination room, and so forth.

Quite a variety of questions is possible in such a standard conventional examination paper. A distinction is made between two main types of questions, namely the more objective type of questions and the so-called own answer, open-ended or more subjective type of questions (Rowntree, 1990: 306,307; Frith & Macintosh, 1988:54,55; Duminy & Söhnge, 1987: 117,125; Mathews, 1985 :104).

By the concept, objective question, is meant questions which require clear-cut answers (Duminy & Söhnge, 1987: 125) and which cannot be influenced by the marker or examiner's subjective preferences (Rowntree, 1990: 307). This type of question can
normally only have a correct or wrong answer and it ensures a uniformity in terms of judgement and marking (Mathews, 1985: 104).

By own-answer, open-ended or subjective question is meant those type of questions in which a short, clear-cut answer is not possible, and in which a learner's answer can differ from that of the examiner (Rowntree, 1990:306; Frith & Macintosh, 1988:55). In such instances, the examiner cannot ensure a complete level of objectivity when marking such answers, and subjective considerations are inevitable. It is important to note here that the standard conventional examination usually consists of both objective and subjective questions.

In order to ensure a clear understanding of these two types of questions, a brief description of the various types will be supplied. A detailed examination of these questions will not be possible within the limitations of this research because that would call for an independent research project.

3.3.2.1 Objective Questions

Under this heading a description of the various objective-type questions will be supplied. Due to the fact that most of the authors who deal with this topic (objective questions) seem to supply a more or less similar list of objective type questions (Bergman, 1981:88-124; Frith & Macintosh, 1988:56-69; Duminy & Söhnge, 1987:125-134; Habeshaw et al, 1993 : 35-42; and Rowntree, 1990: 307-313), no specific references will be made to particular authors.

- Multiple-choice questions appear to be the most familiar and the most frequently used type of objective questions. This kind of question consists of a question (stem) with three of four distractors (wrong or misleading answers) and the actual correct answer. Learners are then required to select the correct answer. White (1989: 78) identifies the following basic assumptions of multiple choice testing, namely:
  - that there are clear right and wrong answers to texts or problems;
that a learner's knowledge about certain subject contents can be portrayed by the brief pieces of writing demanded in such tests;
that a learner's ability to edit correctly will also bear a reliable resemblance to the learner's writing skills.

- The multiple completion questions is merely a variation of the multiple choice questions. In this instance, instead of merely one correct answer that must be chosen, more than one can be correct.

- True-or-false questions. In this case, a statement is made which is either true or false exactly as it is written. Learners must determine which and indicate accordingly on their answer sheet.

- Matching items are a further type of objective questions. It usually consists of two columns (A and B) with corresponding items. When the question paper is set, the different items are supplied in a mixed order and learners must arrange the items in corresponding pairs. To heighten the effectiveness level of such questions, more items should be supplied in the B-column so that learners do not get correct answers as a result of guessing.

- Paired statements: In this instance a number of paired statements are given which can have an effect or influence on each other if a certain third statement is added. Learners must then indicate the correct influence which is determined by the third statement.

- Fill in the blank items: This type of question is created by removing a keyword or words from a sentence or statement and replacing it with a space. Learners must then supply the correct missing word.

- Odd-man-out items: This is another variation of the multiple choice question. In this case learners have to indicate a particular item or statement that does not correlate with the rest of the items or statements in a particular list.
Correct response: This is a further variation of the multiple choice question. In this case, learners must select the correct response to a certain statement.

Assertion-reason items. This is one type of question about which there is no complete unanimity as to its objectivity, because it requires a lot of language comprehension. Here a list containing assertions and possible reasons for these assertions are supplied. The learner must then decide whether each (the assertion and the reason supplied) is true or false. If both are true learners must further decide whether the reason given correctly explains the assertion.

Recall questions. This is another type of objective questions and it differs from most of the other objective-type questions because no clues are provided (i.e. no list of alternatives to choose from). A few variations if this type of questions are the following:

- Short-answer questions: Here learners must formulate a particular answer (usually in one sentence).

- Completion items: (described already).

- Listing of objects: Learners must for example supply a list of the main airports in South Africa.

- Identification questions: In such an instance something such as a diagram or map is supplied from which learners must identify particular items indicated on the diagram or map.

Arrangements items: In such questions the emphasis is more on the relationship between facts or ideas, than on the facts themselves. Learners may for example be required to arrange certain items (places, dates, events) on chronological or distance related or rate of importance basis, and so forth.
• Best answer: This question is a bit more difficult than the others. The wrong answers are not very easy to identify. It can be used to test understanding or interpretation following data or case studies.

It needs to be mentioned that the above list of objective questions might not be complete and no claim is made in this research that all types of objective questions were mentioned. All these types of questions also justify a detailed analysis, which is not possible in this research as stated.

The second main type of question normally found in the standard conventional examination will be described next.

3.3.2.2 Subjective Questions

The most dominant and well-known type of own-answer question, is the essay-type question. Due to its popular use and the fact that it can function as a sole type of examination format, it will be examined separately and in greater detail in section 3.3.3. Other shorter types of own-answer/open-ended questions can normally be distinguished by the keywords or action verbs used.

Examples of such action verbs are:

• Examine: This requires from the examinee to examine matters from different points of view and to identify the advantages and disadvantages thereof.

• Discuss: This verb requires from learners to have insight into a particular matter and to be able to distinguish between applying or assessing knowledge. It also requires an investigation into different aspects of the matter which must then be portrayed in an analytical manner.
• Describe: This type of question does not require any comments or reasoning from the examinee. Rather it requires a presentation of characteristics, facts or results in a logical and well-structured way.

• Analyse: This question requires the examinee to distinguish between the different sections or elements of a particular aspect and to go into a detailed discussion. In this process of analysis, cause-and-effect as well as mutual relationships between variables are established.

• Define: This type of question which is primarily knowledge-based, requires from an examinee to provide a clear, exact and authoritative description of a certain concept or phenomenon in order to convey its meaning effectively to others.

• Evaluate: This question will require learners to assess a certain aspect or phenomenon on the basis of specific criteria. The value evaluation that will eventually be expressed must be adequately substantiated.

• Compare: In this question the requirement is that the examinee should contrast one aspect or concept with another simultaneously. In this process both similarities and differences are identified.

• Explain: The essence of this question is that the examinee should be able to portray information about a certain matter in such a way that the reader can easily understand what it is all about. Aspects that can improve this process are things such as illustrations, descriptions, examples and substantiation of conclusions or results.

• Criticize: This question will require examinees to make a judgement concerning the reliability of certain facts or viewpoints. Such judgements will be based on a thorough exposure and discussion of the positive and negative elements of that particular facts or viewpoints.
• Summarize: The examinee must in this case be able to highlight the main aspects of a particular topic while simultaneously conveying the essential meaning thereof effectively (Van Niekerk, 1995: 13 - 15).

It should be noted that many (if not all) of these action verbs are also used in the essay-type and structured essay-type questions that will be discussed later. In instances where they are used in the standard conventional examination they normally function as short questions.

With this brief descriptions of the various types of objective and subjective questions that can be used in the standard conventional examination, the other types of examinations will be described next.

3.3.3 Essay-type Examinations

Essay-type questions, as mentioned in section 3.3.2.2, are probably the most dominant and relatively independent form of own-answer or subjective questions. Because of its "independence", it is commonly used as the sole type of question in examination papers. Probably the most distinctive characteristic of the essay-type question is the fact that learners are granted the freedom to decide on the exact material that they prefer to include in their responses (Frith & Macintosh, 1988:78).

The use of the essay-type question/examination is not a practice with a recent origin. In fact Duminy & Söhne (1987:117-125) maintain that this type of question dates back to the origin of evaluation itself at round about 2000 BC. Bergman (1981:127) also testifies to this fact.

Habeshaw et al (1993 : 17-31) distinguish between various types of essay questions namely:

• The standard essay question: This type of essay question can take the form of any of the following variations:
- 'Quotation, Discuss or Comment or Query': This essay-type question allows room for learners to challenge different (even expert) opinions.

- 'Write an essay on ...': This is the open-ended type of essay question.

- 'Describe, Give an account of, Compare, Contrast, Explain': The central feature of this type of essay question is that it does not force learners to express their own points of view or conclusions.

- Assess, Analyse, Evaluate: This type of essay question requires the examinee to take an own stance as well as to reach logical conclusions.

- Trick questions: This essay-type question requires learners to really make sure that they understand their work fully so that they are able to "distinguish fact from fiction".

- The role play essay (Simulation): In this type of question, learners have to represent the role or position of another person as if they are that person himself/herself.

- The structured essay: In this essay type question the examiner specifies what contents should be included in the answer. This type of essay question also frequently serve as a sole examination format and it will therefore be discussed as a particular examination type in section 3.3.4.

- Interpretation of evidence: This type of question requires from learners to analyse and interpret information that is provided in the examination question. The advantage thereof, is the fact that it prevents a mere repetition of a prior (to the examination) constructed answer.
• Design questions: This type of question requires from learners to design something new (i.e. a plan, a structure, a policy, or any relevant subject topic). The design question can be coupled with either the interpretation of evidence; role play or the structured question.

• Note-form essay: This type of essay question normally requires a short discussion or explanation of a list of items, concepts or characteristics of some object. It is normally used for the recall of information or for comprehension testing.

• Hypothesis formation: Learners are required to formulate a hypothesis on some topic and afterwards to try and prove that hypothesis. The hypothesis question can be combined with other types of essay questions, for example the interpretation of evidence and the design question.

3.3.4 Structured Essay-Type Questions

The structured essay-type question is also frequently used independently as a sole examination format (as mentioned in section 3.3.2.2) and for that reason it will also be analysed as a separate measurement instrument in this research.

The structured essay-type question plays a very accommodating or compromising role. It helps to bridge the gap between the two extremes, namely the free responses required by the essay-type question and the limited responses required by the objective question. This type of question provides a framework to learners on which to base their answers. Normally a number of headings are supplied in the question focusing learners’ attention on those particular aspects that the examiner wishes to concentrate on (Frith & Macintosh, 1988:69, 70).

Frith & Macintosh (1988:55, 69-78) supply an extensive discussion on the structured essay-type question. Most of the information in this discussion is therefore obtained from them. They describe these questions as basically open-ended in nature but with shorter answers. These questions can be set in the following three ways:
First (the most common use), it is structured in terms of format which can entail the supplying of different section headings in order to make the question more manageable for learners and to focus on specific aspects in relation to the question.

Second, the structure can relate to the material used in the question, for example the editing of a document or quotation; the simplification of a drawing or a diagram, and so forth.

Third, the structure can reside in the particular problem that is posed. In other words, instead of asking learners to describe or discuss certain aspects, a structured problem or dilemma can be posed to learners which they must try to solve.

3.3.5 Written Case Studies

The written case study is apparently a particular format of the written examination which is commonly used in Business Management at a number of tertiary educational institutions in South Africa (see diagram 1 & 2 on pages 302 - 308). Although Priestley (1982: 115) lists the written case study as one of the performance-based types of measurement instruments, it will be discussed in this chapter, because in BM it is already a type of conventional measurement instrument.

In the case of this format, a simulated situation is sketched in the examination paper and most of the questions are based on the case study(ies). The questions normally vary between multiple-choice, short questions, structured essay-type questions and essay-type questions. At TSA on third year level even the open-book format is set on the basis of case studies. The main similarity between this type of examination and the standard conventional examination is the fact that the learners do not have any prior knowledge about the exact contents of the examination paper.
3.3.6 The Open-book Examination

Although not yet in common use on a wide spectrum, the open-book examination is already a well-known educational measurement instrument and is propagated by a number of authors (i.e. White, 1983:132; Bergman, 1981 : 82, 83; Frith & Macintosh, 1988 : 79, 80; Habeshaw et al, 1993 :67). The concept, open-book examination, refers to an examination in which learners are allowed to use and refer to textbooks, notes and other sources during the course of the examination. Learners are thus allowed to use their textbooks in order to answer the question(s).

It becomes evident that in such an examination very little attention is paid to the retention of facts by learners which would imply the use of lower cognitive skills such as knowledge. Instead it would appear that the focus is rather on higher cognitive skills. White (1983:132) testifies to this assertion because he regards the open-book examination as one of the most effective ways of assessing the higher levels of cognitive ability of learners.

In conclusion it is useful to note that it seems as if the open-book examination is in increasing use in universities and colleges internationally (White, 1983:132). The open-book examination is in fact a move away from the standard conventional examination and can in fact be regarded as an alternative to the standard conventional examination. However, it still remains a written examination and the focus in this research is to find alternatives to the written examination.

3.3.7 The Pre-Set Question (Seen Examination)

The pre-set question (White, 1983:132; Frith & Macintosh, 1988:79) or seen examination (Habeshaw et al, 1993 : 65) is in fact only a variation of the essay-type question. The only difference lies in the fact that learners in the case of such a pre-set question(s), will have prior insight or knowledge as to the nature and wording of the particular question(s) that will appear in the examination paper. This will give learners the opportunity to carefully ponder over the question topic, to choose a line of attack
and to prepare their argument(s) properly. It is very evident that such a type of examination will focus on the learner's higher mental qualities. Habeshaw et al (1993: 65, 66) regard this form of examination as a more realistic way of testing because it in fact evaluates learners' ability to undertake research and to use resources effectively.

Habeshaw et al (1993: 65, 66) distinguish between two types of seen examinations, namely:

- The nine-month examination: In this case the lecturer provides the learners with the examination questions right at the beginning of the course. This approach can help to structure the year's work and to identify the most important aspects. The main disadvantage is the fact that learners might only be concerned with the aspects that will be examined and neglect the other parts of the work which might also be important. Not all the questions provided in the beginning will necessarily have to appear in the examination.

- The one week examination: In this case the examination paper is made known to learners only one week before the examination: Learners can then use notes, and other sources to prepare adequately.

With these brief descriptions of the main and most familiar types of written measurement instruments, it seems fitting at this stage to illustrate the global use of these measurement instruments. Not all of these measurement instruments might be mentioned in the following section, but the authors who discuss them are adamant that those instruments are already in common use.

3.4 CONVENTIONAL SUMMATIVE MEASUREMENT INSTRUMENTS - A GLOBAL VIEW

The intention with this section is to give an illustration of probably the most common and popularly used summative measurement instruments in educational institutions
across the globe. This global contextualization of summative educational measurement instruments is very important and even crucial, because it will help to compare or judge the relevance, applicability and suitability of those measurement instruments used at educational institution in South Africa and in particular those used at TSA for Business Management.

The reasons why examples of the various admission examinations for tertiary/higher education are recorded here, are the following:

• It was difficult to get hold of first-hand information regarding the evaluation and measurement instruments of different tertiary educational institutions world-wide and even if such information could be obtained from one or two institutions per country, it would still not give a representative picture of measurement instruments used in those countries. A quest to describe the measurement instruments used at all universities or technikons in all countries in the world would have been an immense task and it would have hampered the focus of this research.

• No extensive research has seemingly yet been undertaken to compare the examination types and formats of tertiary educational institutions globally as is the case with these tertiary admission examinations.

• The sources consulted give a comprehensive overview of various forms of summative measurement instruments used in different countries for educational purposes on a national basis.

• Certain authors state that many of the same tests and examinations used for admission to tertiary education (and even the ones used in secondary education, are also used in higher education institutions (Courts & McInerney, 1993:21). So the nature and format of these admission examinations provide a fairly representative picture of some of the forms of measurement used at (tertiary) educational institutions globally.
Courts & McInerney (1993:20) for example provide some illustrations of tests used by the Educational Testing Service (ETS) for admission to tertiary education in the United States that are also used at tertiary educational institutions in the country:

- The American College Testing programme (ACT) which is used to make decisions about college admission;
- CLEP (College Level Examination Program) which is used for the granting of academic credits by means of examinations;
- The GRE (Graduate Record Examination) used for admissions to graduate school;
- The PPST (Pre-Professional Skills Test) used to assess learners' growth and development as a result of instruction.

Following this brief introduction the more older or more conventional instruments in common use will be discussed next.

3.4.1 Older forms

It is important to note here that although a number of the summative measurement instruments discussed here are called tests, they do in fact function on a summative level, because they serve as a measure to determine how much knowledge and skills learners have obtained or mastered in the course of their secondary school career (in other words at the end of their secondary school career).

3.4.1.1 The United States of America (USA)

Noah & Eckstein (1992: 16 - 18; 1990:86) state that no official national examination has to be written by learners at completion of secondary school and for entrance to higher education in the USA. Requirements for graduation from secondary school are determined by the educational authorities in each of the 50 states. Only the states of New York and California administer state-wide achievement tests for the purpose of awarding a graduation certificate from secondary school. In most of the remaining
states the conditions for the achievement of a high school diploma is the completion of a set number of courses without any final examination therein.

Noah & Eckstein (1992:17) do however mention that there are a number of private organizations (called testing agencies) who provide certain tests in individual subjects. The first of these is the Educational Testing Service (ETS) which administers the Scholastic Aptitude Test (SAT), the Achievement Test (AT) and the Advanced Placement Test (APT). The second organisation is the American College Testing Program (ACTP) which administers the American College Test (ACT).

ETS was established in 1947 by the College Entrance Examinations Board (CEEB), the American Council on Education and the Carnegie Foundation for the advancement of teaching, because many of their functions were overlapping (Solomon, 1988:55). ETS assumed responsibility for all three organisations' testing services. ACTP was established in 1959 and it administered admission tests mostly for public universities in the central southern regions of the US (Solomon, 1988:57).

The tests administered by these testing agencies are mainly multiple-choice in nature and computer-scored (Solomon, 1988:65). Because of the growing criticism against the multiple-choice format some alterations to the formats of the ACT and the SAT were made to include essay type questions (Noah & Eckstein, 1992: 16 - 18). Most colleges and universities in the US require learners to undergo such tests in addition to their high school work, for admission purposes.

According to Gipps (1990:59) the National Evaluation of Educational Progress (NAEP), which was transferred to the Educational Testing Service in 1983, also mainly makes use of the multiple-choice format in their tests and examinations. However, they have received funds from the National Science Foundation in the US to develop and produce tests and examinations of a more practical nature in order to evaluate the higher order skills in mathematics and science (Gipps, 1990 : 60).
The NAEP also developed a reading scale which was intended to measure learners' proficiency in reading. The scale that "goes from 0-500" includes five proficiency levels, namely:

- The rudimentary levels: 150
- The basic level: 200
- The intermediate level: 250
- The adept level: 300
- The advanced level: 350

This scale is based on the item-response theory (Gipps, 1990:60).

With regards to the various institutions of higher education it is notable to find that the types of examinations and tests used by the various institutions, vary remarkably even within the same institution. Jordan makes the example of the University of Tennessee at Knoxville, where two departments make exclusive use of essay-type examinations, while four other departments only use the multiple-choice format (Jordan, 1989:115). Jordan further states that there is a tendency in the US that where there are very large numbers of learners involved, institutions tend to prefer the multiple choice format because it is machine-scorable (1989:116).

Oral examinations according to Jordan (1989:118) are also frequently used in many universities and especially at postgraduate level. Such oral examinations are normally used as follow-ups on the written examinations in cases where learners' answers were a bit ambiguous or not clear enough.

There are also a number of what Jordan calls, commercial tests/examinations, which are administered at various higher institutions (1989:122). Examples of these are:

- The Graduate Record Examination (GRE) (which was developed by the Graduate Record Examinations Board) is a combination of tests focusing on aptitude and achievement and serving as the gateway to the graduate school's degree
programmes. The GRE aims at the evaluation of different cognitive levels, i.e. memory, comprehension, analytic thinking, evaluation and so forth.

- The Academic Profile (which was developed by ETS is a combination of tests which focus on content and outcomes in three areas, humanities, social science and natural sciences. The functions that are examined in each content field are:
  - college level reading
  - college level writing
  - critical thinking and
  - using mathematical data (Jordan, 1989:122)

- The College Outcomes Measurement Program (COMP) (which was developed by the American College Testing Program). This is a four hour long test which consists of fifteen activities which are taken from stimulus materials in the adult world (Jordan, 1989:123).

According to Courts & McInerney, in the American higher education, there was especially since the middle 1980's a strong external (government, social & community institutions) pressure for the introduction and implementation of effective forms of evaluation of learners' academic progress and achievement (Courts & McInerney, 1993:16).

In the case of the Northeast Missouri State University (NMSU) for example, external pressures demanded the use of certain standardized tests for evaluation purposes (Courts & McInerney, 1993:18). In trying to meet those demands, the NMSU introduced the following measures:

- A "value-added" type of measurement system that included the use of the ACT and the ACT COMP;
- The use of a national examination such as the Graduate Records Examination (GRE) field examination;
- A number of surveys by means of questionnaires;
• Numerous other national examinations such as the ETS placement examination, the ACT COMP writing examination, the ACT COMP speaking examination and so forth (Courts & McInerney, 1993:18).

The NMSU practice confirms the fact that in some higher education institutions, the same measurement instruments used at secondary school and for tertiary admission purposes are frequently used to assess learners' academic progress as mentioned in section 3.4.

Higher educational institutions in Florida have also been subjected to such external pressures. This led to the introduction and use of the Florida College-Level Academic Skills Test (CLAST). This test was aimed at ensuring that every learner that is allowed to the senior phases of college education, has mastered certain basic skills in the following areas: reading, writing and quantification. Many people however regard this test (CLAST) as a type of "gatekeeping" examination because it was in essence used to ensure basic academic competencies in all college graduates (Courts & McInerney, 1993:20).

Heywood states that about 250 institutions in the US use the College Outcomes Measures Project (COMP), developed by the American College Testing Program (ACT). This measurement project is used to assess:

• general knowledge;
• problem-solving skills;
• communication skills;
• the clarification of values (Heywood, 1989:14)

Fisher & King (1995:5) state that standardized testing is a predominant component of American evaluation and it remains a prerequisite for admission to higher education. Quite a number of colleges in the US use standardized tests for the examination of both general and specialized knowledge (Heywood, 1989:14).
In conclusion it would appear that although various examination formats are being used in the US at different institutions and agencies, the most popular examination format or measurement instrument is the multiple-choice question. It is no wonder therefore that Dwyer (1990:24) states that the US can be regarded as the “bastion” of multiple-choice testing and examinations. As a result of the extensive use of multiple-choice forms of evaluation, technology in the form of optical scanners was also utilized to smoothen the process (Dwyer, 1990: 25).

3.4.1.2 England and Wales

Despite the fact that universities and colleges in England and Wales use their own criteria to select learners for enrolment and that the final secondary school examination is not regarded as sufficient enough for admission into higher education, the UK also uses national evaluation (Noah & Eckstein, 1992:12).

The main body responsible for this national evaluation is called the Evaluation of Performance Unit (APU) (Gipps, 1990:53). The APU, which functioned as a unit within the Department of Education and Science, see as their main task the development and establishment of educational standards (Gipps, 1990:54). The format of some of their examinations is the essay-type (Noah & Eckstein, 1990:95).

Between the period 1950 - 1988 the main formal school examination was known as the General Certificate of Education (GCE). Two types existed, namely the GCE Ordinary level and the GCE Advanced level. The General Certificate of Secondary Education (GCSE) was later introduced to integrate the predominantly university orientated school-leaving examination (GCE) with the more general examination, the Certificate of Secondary Education (CSE) (Noah & Eckstein, 1992:13).

Reddaway (1988:28) state that the nature of the GCE was achievement tests which intended to measure learner ability in various subjects and it was mainly administered by the University of Cambridge Local Examinations Syndicate (UCLES). UCLES makes use of various measurement techniques which include:
- essay-type questions;
- structured essay-type questions;
- practical examinations;
- Coursework;
- teacher evaluations;
- tests; and
- multiple-choice questions (Reddaway, 1988: 29).

It is also of interest to note that in the GCE advance level technology examinations, UCLES, makes use of the open-ended essay-type question (Reddaway, 1988:30).

In conclusion it would appear that in contrast to the USA where multiple-choice questions is the dominant format of measurement, the essay-type question still features strongly in the British educational system.

3.4.1.3 France

In France the examination system is very centralized, even more than in the case of Japan (Noah & Eckstein, 1992:7 - 9). The school-finishing examination and university entrance examination is called the "baccalaureate". Successful completion of this baccalaureate entitles every learner entrance into any French university. The only areas in which some further requirements are set is in the field of medicine, dentistry and engineering (Noah & Eckstein, 1992: 8).

The format of the examinations written for this baccalaureate are the following:

- standardized forms of examination;
- long essay-type questions;
- mathematical computations;
It appears at the stage of Noah & Eckstein's research that the multiple-choice format has not yet taken central stage in France.

3.4.1.4 Federal Republic of Germany

Noah & Eckstein (1992:9) state that examinations in Germany have always been largely regionally and locally controlled. The examinations are also locally set. The certificate obtained after the final school examination is known as the Abitur.

The older tradition in Germany used oral examinations and although it was replaced by the written examination, they still use the oral examination to a limited extent (Noah & Eckstein, 1990:90; 1992:9 - 11). Currently the format of the examinations for the abitur is also long essay-type and mathematical computations (Noah & Eckstein, 1992:17).

3.4.1.5 Sweden

The Swedish educational system uses the process of continuous evaluation (Marklund, 1988:99).

Sometime in the 1970's Sweden abandoned the practice of final secondary school examinations and replaced it by a system in which final marks are based on a combination of marks obtained for class exercises and homework as well as in the various tests conducted during a pupil's school career (Noah & Eckstein, 1990:93). The reasons for this dramatic swift to continuous evaluation are given as:

- to decrease pupil strain and anxiety;
- to ensure more trustworthy predictions learners' of university success;
- to eliminate or prevent socio-educational inequities in the evaluation process;
- to improve or promote the diagnostic and predictive value of tests;
- to make it possible for teachers to set national criteria against which to evaluate their own work.
Sweden does not require a special entrance examination for admission to higher education. Rather tertiary institutions evaluate the average marks obtained by learners at secondary school. Admission to higher education became more based on work experience as well as the level of maturity and other credentials (Marklund, 1988:104).

However, Sweden has not completely abandoned examinations. It is merely that, since 1988, the tests and examinations were to be complemented by continuous evaluation (Noah & Eckstein, 1990:94).

Some of the formats used in Swedish final school examinations are the following: Standardized tests which measure skills and knowledge in different subjects. These tests are both achievement-based or diagnostic-based. The achievement tests are based on the multiple-choice format and are administered by teachers themselves (Marklund, 1988:94, 95).

3.4.1.6 Union of Soviet Socialist Republics

According to Noah & Eckstein (1992:11, 12) the certificate at the end of secondary school in the Union of Soviet Socialist Republics is known as the “attestat zrelosti”. The examinations for this certificate are mainly locally administered but there is also a strong centralized control. Similar to the process in Japan, aspirant tertiary learners have to write two examinations.

In terms of the format of the examinations, they continued the German tradition of oral examinations in addition to written examinations. They also use the essay-type questions and mathematical computations (Noah & Eckstein, 1992:17).
3.4.1.7 The People’s Republic of China

Noah & Eckstein (1992:15) state that China can be regarded as the “birthplace” of examinations. The Imperial Examination System which was started in AD 909, lasted until 1909 during the overthrow of the Qing Dynasty. Examinations still served as a gateway to tertiary education until 1966 after the Cultural Revolution. Thereafter admission to tertiary education was determined on the basis of admission tests, class background, work experience and “political reliability”. However, the examination system as vehicle to senior secondary and higher education was reintroduced by the Chinese government in recent years (Noah & Eckstein, 1992:16).

The examinations in the Republic of China are also very highly competitive, similar to those in Japan (Noah & Eckstein, 1990:91). Traditionally they used the essay-type question, but in recent years they also switched to the multiple-choice format, but coupled with the short question type. Although they haven’t yet started with computer-scoring of the multiple-choice questions, there appears to be a move in that direction (Noah & Eckstein, 1990:91, 92).

3.4.1.8 Japan

In 1984 Japan had a total of 460 universities of which 95 were national ones, 34 local public ones and 331 private ones. In terms of junior colleges the number stood at 448 and in terms of technical colleges, at 62. What is noteworthy about the popularity of these institutions is the fact that the largest majority of learners at universities in Japan in fact attend private universities (76 %) whilst the minority (21 %) attend national universities and only 3 % attend the local public universities (Hidano, 1988:9). Based on this information it would seem possible that the private universities might have a very big influence on higher education in Japan.

Japan also makes use of admission tests for entrance into tertiary education. The first test used for university entrance for learners after World War II, was the scholastic
aptitude test which was constructed by the National Institute for Educational Research. This test system was however terminated in 1954 (Hidano, 1988:11).

In 1963 an institution by the name of the Talent Development Institute, was established. This institute developed different types of tests for the purposes of university selection, namely the scholastic aptitude test, the vocational aptitude test and the achievement test. This institute was closed in 1968 because too few universities utilised their services (Hidano, 1988:12).

In 1976 the National University Association decided on a two-staged test approach. In the first stage all national universities administered the same achievement test, while at the second stage each national university administered its own admission examination (Hidano 1988:12).

The test at the first stage became known as the Joint First Stage Achievement Test (JFSAT). This test focused on applicants’ attainment in terms of basic and general knowledge and skills which were obtained during senior secondary school. The JFSAT is written on a national basis and administered on a centralized basis by the government (Noah & Eckstein, 1992:7). Since 1977 the JFSAT was administered by the National Centre for University Entrance Examinations (NCUEE) (Hidano, 1988:12).

The second-stage examination focused on those skills and abilities that each particular university regards as essential for admission at their institution (Hidano 1988:12). Criticism against these university entrance examinations is that it is too factually-based and competition oriented, while requiring the recall of isolated parts of information without real understanding on the side of learners. These entrance examinations have frequently been termed “examination hell” (Frost, 1992:25).

The format of the JFSAT is the multiple-choice type (Noah & Eckstein, 1990:87). The subject areas evaluated in this test are: Japan language, social studies, mathematics,
science and foreign languages (Hidano, 1988:18; Frost, 1992:28, 29). It appears as if the second stage examination administered by the universities themselves, mainly consisted of the essay-type format, because Hidano (1988:24) claims that the essay-type examination (despite the increasing use of the computer-scored test) still remains the most popular format in many universities in Japan.

3.4.1.9 Australia

Australian schools used a two-level form of examination until the middle of the 1960’s. The nature of the examination is as follows:

- After ten years of schooling an examination was written which granted the examinee entrance to employment, apprenticeships and other forms of training.

- After a further two years of schooling, another examination was written which granted admission to tertiary education (Keeves, 1988:71).

Some of the various Australian education systems use the following examination/testing formats:

- The Australian Scholastic Aptitude Test (ASAT);
- Reference tests;
- Multiple-choice questions (Keeves, 1988:82, 83).

Afressa (1995:287) however states that in public examinations the format of most of the examination papers is the essay-type although the objective type question is also frequently used.

3.4.1.10 Africa

In Africa there is also the use of public examinations which are administered by an
agency outside the school (i.e. a national examinations council or the education ministry). Kellaghan (1992: 98, 99) distinguishes three such examinations, namely:

- The first at the end of primary school;
- The second after two or three years secondary education;
- The third at the end of secondary school.

In terms of the nature and format of these examinations, Kellaghan states that they are almost exact copies of the European types of examinations (1992:99). The reason for this is probably the long colonial influence of Europe in Africa as well as continued financial assistance to many African countries. Kellaghan for example states that the University of Cambridge's Local Examinations Syndicate set the school leaving examination of about seven African countries up to 1981 (1992:99). So, although countries in Africa formed their own examination boards, they did not drastically break away from the European models of examinations.

The examination systems of Ethiopia, Lesotho, Malawi, Swaziland and Zambia were mainly based on the English model but also on the American examination system to a limited extent (Kellaghan, 1990:100). Some general characteristics of the examinations of these countries are the following:

- Formal
- terminal
- subject-based
- external to the school and administered by the Ministry of Education or Examination Councils
- A great emphasis on written work with a neglect of practical, oral or course-work assessment.
- The main format of the examinations was the essay-type question, but it was in some cases partly and in others fully replaced by the multiple-choice format.
- All five the above countries use external examinations at various stages in the school career (end of primary school, end of the junior cycle of secondary school
A number of countries in Africa switched to the use of continuous evaluation in the place of or additional to final examinations. Some examples of countries who changed to continuous evaluation are: Tanzania, Papua New Guinea and Nigeria (Pennycuick, 1990:107 - 109).

Tanzania introduced the system of continuous evaluation in 1976 (Pennycuick, 1990:107). The main aims for this shift were:

- To get rid of examinations that is merely such as an “ambush” to learners;
- To reduce the great emphasis that was always placed on written examinations.

The continuous evaluation accounts for 50% of a learner’s final mark while the formal examination account for the remaining 50% (Pennycuick, 1990:107, 108).

Papua New Guinea started using the system of continuous evaluation since 1975 in combination with a reference test, called "the mid-year rating examination “(MYRE) (Pennycuick, 1990:108). The MYRE system functioned on the basis of awarding distinctions, credits and pass marks according to an internal rank order. It was also more skills-based than merely content-based. However, the MYRE system was later abandoned in 1981 and replaced by syllabus-based examinations at the end of the school year. These syllabus-based examinations counted 50 of the final mark while the remaining 50 was based on continuous evaluation (Pennycuick, 1990:108).

3.4.1.11 The Seychelles

Secondary education in the Seychelles also use continuous evaluation in their National Youth Science programme (which is a national centralized and uniform secondary education system) (Pennycuick, 1990:109). The characteristics of their continuous evaluation are the following:
• objective-based;
• administered by subject teachers;
• continuous observation;
• work folios;
• oral or written tests; and
• homework.

All of these measures contribute towards the cumulative learner profiles (Pennycuick, 1990:109).

3.4.1.12 South Africa

Not much will be recorded about South African evaluation practices here because more exact details will be provided in section 3.5. The only aspect of relevance that needs to be mentioned at this stage with regards to admission examinations for tertiary education, is that the body mainly responsible for admission examinations in South Africa, is the Joint Matriculation Board (JMB) (Shaw, 1977:84). The JMB was established by law in 1916 (1977:84). The main types of examinations in South African schools administered under the control of the JMB were written examinations, orals and practical examinations in certain subjects (Shaw, 1977:95 - 98).

3.4.1.13 Conclusion

Although the information obtained from this comparative description will not be used for statistical analysis, it is possible to make some observations based thereupon. These observations are the following:

• It does appear from this comparative illustrations, that the two main written measurement formats world-wide are the essay-type and the multiple-choice questions.
• It also became clear that the multiple-choice question is in increasing popularity world-wide and that is has spread across the globe already. However, despite its growing popularity and phenomenal growth, the multiple-choice question has not succeeded in eliminating the essay-type question. In fact, it would appear as if in some countries the essay-type question is preferred above the multiple-choice. Reasons for this tendency will probably cleared in sections 3.5.2 and 3.5.4 when a brief scrutiny of these two measurement instruments is supplied.

• The fact that a number of countries have moved away from summative evaluation to exclusive formative (continuous) evaluation.

• In most countries a variety of measurement instruments are used. This is in fact correlating with the advice of various authors as mentioned, that rather a variety or multifaceted type of evaluation/measurement be used.

Following these comments the more recent measurement tendencies and formats will be described next.

3.4.2 New developments

A number of new developments have taken place on the educational evaluation and measurement scene. In chapter 2 (section 2.4) illustrations were already given of the new directions in scientific thinking and how those movements also have an impact on education as a science. In this section a brief exposition of some of the important developments with regard to educational measurement will be supplied. Three aspects will be discussed, namely some suggestions and recommendations by major scholars in the field of educational measurement; aspects that should be taken into account in the process of developing new measurement instruments for the new context that we live in; and finally, some examples of such new developments in some tertiary institutions world-wide.
3.4.2.1 Suggestions/Recommendations

A number of scholars in the field of educational evaluation have made some very useful suggestions and the view in this research is that those suggestions should be seriously taken account of by all educators. Here follows a brief summary of what some of them say.

Justiz & Kameen (1989:40, 41) suggest that, in order to address the different needs of multicultural learner populations, college evaluation programmes should use multiple measures which will help construct accurate profiles of learners' academic progress and performance. Evaluation programmes should also include aspects such as diagnostic feedback, remediation and counselling which can be used to the benefit and assistance of minority learners. Such multiple measures of evaluation can include the following (which can all be administered to the same group of learners alternatively):

- essays
- orals
- portfolio and performance evaluations,
- simulations,
- seminars, and so forth. (Justiz & Kameen, 1989:41)

Gipps (1994:25,26) states that due to the world-wide changes in technology, communications and economic principles, there is a growing quest towards understanding and logical thinking in all spheres of life (workplace, education, society etc.). In terms of education, there is a call for educational programmes that will train and prepare learners to reason and think logically in order to be able to address the challenges and problems that they will have to face. The emphasis is therefore increasingly on problem-solving skills and logical reasoning. Gipps (1994:26) argues that the acquisition of higher order skills such as logical interpretations, logical arguments, problem-solving skill, etc., should be developed from a very early age in learners and that the evaluation processes should be adapted accordingly.
Fisher & King (1995:6) also emphasize the increasing demand for training of learners in logical reasoning, thinking and problem-solving skills, and so forth. Other skills that learners should be adequately trained in are:

- the ability to work with continuously changing data;
- decision-making skills;
- skills in teamwork

Fisher & King (1995:7) state that educational institutions should stop measuring the deficiencies that learners have, and rather focus on the skills that they have mastered. Evaluation should therefore serve as opportunities to learn.

Heywood (1989:19) also emphasizes the need for evaluation programmes that use multiple measures to determine learner academic progress and achievement.

Courts & McInerney (1993:22) state that there is already a move away from standardized examinations towards qualitative evaluation methods. Courts & McInerney (1993:25) also argue that a modern demand placed an evaluation practices is that it should became more “learner-centred”. In other words, the evaluation methods should try to establish:

- what the learners are actually learning;
- why they care about what they learn;
- what they will in fact wish/prefer to learn;
- they own views on personal effective learning, and so forth.

Mathews (1985:55) states that there is a growing demand that educational institutions should be more concerned with other human attributes, such as depth of judgement, logical perspectives, understanding human motivation, and so forth, than merely with the intellectual abilities of learners.
Brown & Knight warn against the use of a single method of measurement or evaluation of learner academic progress and achievement (1994:44). Brown & Knight (1994:45) further state that measurement methods will have to change in most tertiary institutions in England because of the almost compulsory need for change in the curricula. Some of the main compelling factors for such change in the nature of curricula world-wide according to them are the following:

- First, the growing need for educational programmes that will advance a greater collection of achievements in learners with regard to what they have learned;
- Second, the ability of institutions of higher learning to handle the vast growth in learner populations, without having to increase staff numbers accordingly. In other words, cheaper ways of teaching have to be developed (Brown & Knight, 1994:45).

Brown & Knight (1994:45) therefore argue that the new forms of evaluation should rather aim towards the empowerment of learners.

There are probably many more recommendations put forward by various scholars, but the ones listed above give a good indication of the main feelings among educators and evaluators. The next section will therefore be used to illustrate some crucial factors that must be kept in mind in all future planning with regards to educational evaluation/measurement on tertiary level, and especially in South Africa.

3.4.2.2 Aspects to consider

There appears to be a growing emphasis in educational evaluation circles on the prevention of bias in examination papers and questions. Matthews (1989:20) for example states that a deliberate attempt was launched by the Texas Academic Skills Program (TASP) to eliminate and prevent all forms of bias in tests used by them. A panel was even formed for this purpose. Texas educators were at a later stage even provided with a manual on bias prevention in tests and examinations (Matthews, 1989:21).
Also in England special attempts are made to eliminate bias from examination questions. The School Council tried to develop ways in which to prevent or minimize cultural or ethnic bias in examination papers by using the following criteria on which examination questions should be scrutinized:

- Establishing whether there is a variety of evaluation methods (for example tests, coursework, and so forth) which will allow learners to portray skills or interests that are particularly common to their own culture (i.e. intellectual-cognitive versus technical-practical).
- Whether the language usage of questions will minimize or maximize difficulty in understanding the questions.
- Whether examiners are aware of and sensitive to the difficulties that might be experienced by second language users in answering the questions.
- Is the marking criteria open or susceptible to various culturally inspired ways of expression or does it only demand one cultural form of expression? (Mathews, 1985:134)

Justiz & Kameen (1989:38) also mention the growing emphasis on the prevention of bias in tests and examinations in the American higher education, because of the growing multicultural and multilingual composition of the American society.

Other factors that should be taken account of in the development of new measurement instruments is the numerous shortcomings in existing evaluation/measurement instruments. Frazer & Nieman (1995:127, 128) for example mention a number of shortcomings with regards to evaluation practices in most distance education institutions in South Africa. These are:

- No real positive changes or developments in evaluation methods with regards to distance education for the past two decades;
- Still a high focus on the evaluation of lower cognitive levels of achievement (knowledge and comprehension);
• Too limited scope of open-ended questions in examination papers;
• A great lack of diversity in terms of evaluation policies and methods.

These shortcomings in existing evaluation/measurement methods are very relevant with regards to TSA because of the fact that it is also a distance education institution. They should therefore be taken into account in the development of new and more suitable measurement instruments.

3.4.2.3 Examples of new approaches to measurement

There appears to be a move towards portfolio evaluation in the USA. White (1989:80) states that a number of tertiary educational institutions in the US is already using the portfolio method of evaluation. These institutions are:

• the University of Minnesota;
• the State University of New York; and even
• the Educational Testing Service (who is supporting research into the use of portfolio evaluation).

White (1989:80) further states that portfolio evaluation is probably one of the main forms of evaluation according to the process-model of evaluation referred to in chapter one (section 1.3).

There also seems to be a major shift towards the use of performance evaluation methods in general (including portfolios, profiles and records of achievement in the US as well as in the UK (Gipps, 1994: 108 - 112; Assiter & Shaw, 1993: 19 - 21; Broadfoot, 1986: 1 - 11; Hall, 1989:14 - 17).

Sims (1992:38) provides another example of new developments with regards to new evaluation methods that started in the 1970's already. He describes Alverno College (a small private women's' college in Milwaukee, US) where an evaluation centre was established in 1973. This evaluation centre uses a multiple measure evaluation
program in which the aim is to establish whether learners have mastered the following skills:

- critical thinking;
- problem solving;
- communicating;
- making value decisions from the existing curriculum (Sims, 1992:37).

The format of these measurement instruments include: situational exercises; job simulations such as business games, discussion groups, reports, presentations, and so forth (Sims, 1992:38).

Another new development in terms of different evaluation formats and approaches that started in the 1970's was at Northeast Missouri State University (NMSU). NMSU introduced a “multifaceted, value-added” evaluation programme (VAP) in 1974. The VAP makes use of an external examination which is produced by the American College Testing Program (ACT). This examination is first administered as a pre-test to learners intending to enrol at the university, and after two years of study at the university, it is administered as a post-test to the same learners. The difference between the scores of the pre-test and the post-test gives an indication of the value that was added to every individual learner (McClain, Krueger & Taylor, 1989:35).

The VAP aims to measure achievements and developments in the person as a whole. This includes aspects such as: knowledge (cognitive); analytic ability; skills; values; cultural awareness, and so forth (McClain et al, 1989:33). This programme focuses on the evaluation of changes within individuals after a period of educational instruction. The intention is to determine the value university instruction adds to learners as they proceed through their studies. This programme required the keeping of records of accountability for the growth levels achieved in terms of the individual learner, the educational programme and the institution in its totality (McClain et al, 1989:34). The VAP is essentially a performance evaluation plan and focuses on outcomes.
The VAP is to a certain extent based on the principles of the value-added tax systems used in a number of countries. According to the VAT system, every enterprise must account for the value it adds to consumer goods produced by it. In order to determine the real "growth" in the product's value, such an enterprise has to assess the actual value it added to the product (McClain et al, 1989:34).

These are but a few examples of some new or alternative approaches to the conventional written examination, but it shows that there is a move away from traditional forms of evaluation/measurement.

3.4.2.4 Conclusion

A number of observations can be made on the basis of the different views and arguments of the various authors in sections 3.4.2.1 - 3.4.2.3.

- First, it becomes evident that there are strong recommendations towards the use of a variety or multiple forms of measurement/evaluation for the sake of different individual learners as well as multicultural and multilingual learner populations.
- The development of measurement/evaluation instruments that will prevent any bias against or for any particular cultural group in a multicultural society.
- There appears to be a need for the elimination of shortcomings in conventional and existing measurement instruments.
- It appears that there is indeed already a move away from the mere evaluation of knowledge towards the evaluation of higher cognitive skills.
- The emphasis seems to fall on performance and the use of practical and real world skills.

With this brief summary of some current trends in the summative evaluation context, a brief description of the South African situation with regards to measurement instruments used in Business Management at tertiary educational institutions will be supplied.
3.5 A SCRUTINY OF CONVENTIONAL MEASUREMENT INSTRUMENTS

A scrutiny of the various summative measurement instruments is necessary in order to discover why there is such a lot of criticism against the conventional measurement instruments as mentioned in chapter 1 (section 1.3). Such an evaluation will also make it clear to the reader why it is felt that there might be a need for alternative summative measurement instruments.

The criteria against which these variations will be examined, are the following:

- **Advantages:** In this case an attempt will be made to determine all the positive elements in a particular measurement instrument that afford credit to that instrument.

- **Disadvantages:** The main shortcomings and areas of criticism will be identified.

- **Validity:** This is one of the basic criteria that any measurement instrument must comply with. Duminy & Söhne (1987: 165) describe validity as meaning that a certain measurement instrument is able to measure what it is supposed or intended to measure.

- **Reliability:** Reliability on the other hand refers to the fact that a particular measurement instrument is capable of rendering similar results on a continual basis (Duminy & Söhne, 1987: 164).

- **Suitability for the TSA:** The intention will be to determine to what extent each of these measurement instruments comply with the unique conditions of TSA as a tertiary, technical, distance education institution in South Africa. These conditions were illustrated in chapter 1 (section 1.3) but here is a summary of those conditions:
  - The aspect of distance education.
  - The aspect of self-study.
  - The practical nature.
- The aim of the mastery of course material and skills.
- The aspect of second and/or third language learners.
- The aspect of multiculturalism.

3.5.1 Scrutiny of the standard conventional examination in general

3.5.1.1 Advantages of the standard conventional examination

Cawood et al (1988:123-126) list a number of advantages of the standard conventional examination in as far as the learner, the lecturer and society in general are concerned. Only those advantages that are regarded as relevant to this research will be mentioned. Here follows a short summary of those advantages:

- The results of such (conventional) examinations serve as an indication to both learners and lecturers of what amount of progress learners have achieved for a particular year or period of study.

- Examinations serve as a strong motivating factor to both learners and lecturers to work harder and strive for optimum achievements.

- This motivation also enables both learner and lecturer to be successful in their particular endeavours.

- The coercive element in examinations forces learners to get the study material under control and that in turn enables them to be able to retain and utilize that knowledge.

- Examinations also help learners to develop certain skills, for example how to complete a particular task in a prescribed period of time; how to work under pressure, and so forth.
- Examinations further serve as a means or basis to determine whether a learner is fit and capable to be promoted to a higher/further year of study.

- Examination results also serve as a criteria for the awarding of certificates, diplomas or degrees for employment and other purposes.

Mathews, 1985:20-58; 105 - 136; 185) also identifies a number of advantages of the standard conventional examination. In addition to the advantages mentioned by Cawood et al, he mentions the following:

- Mathews (1985 :20) argues that examinations are a very cost-effective form of evaluation, especially if administered to large numbers of learners. Also in terms of the marking process it remains a relatively cheap method.

- Second, he asserts that examinations were refined during its long history and in the process has attained a level of reliability that cannot easily be surpassed by another form of evaluation (1985:21).

- Mathews also argues that examinations can serve a formative function in education, provided it is administered more than once per year/semester (1985:29).

- Mathews sees as a further advantage of examinations the fact that examination questions can be stored and re-used on numerous occasions (1985:113).

There might be other possible advantages of the standard conventional examination, but the list supplied above does include many of those most generally regarded as the main advantages of examinations. Although some of these advantages might indeed be valid and relevant, there seems to be much room for criticism. It can be mentioned for example that standard conventional examinations are not necessarily the best or most effective way to indicate the progress of learners as will become evident in the course of this research, nor is it the most effective motivation factor for learners and lecturers. As a result of that “coercive” factor of examinations as mentioned, learners tend to
"cram" a lot of information merely to pass the examination, but forget most of that information and material very shortly after the examination. Also the assertion that examinations promote the acquirement of certain skills is not very valid, because precisely due to the fact that the average learner experiences a tremendous amount of stress and tension while writing the particular examination, such learners are not optimally capable of functioning and employing their skills effectively.

Following these few comments concerning the advantages of the standard conventional examination, the actual disadvantages of this instrument will be examined next.

3.5.1.2 Disadvantages of the standard conventional examination

Indications seem to show that the disadvantages of the standard conventional examination actually outnumber the advantages thereof. A number of authors have identified long lists of disadvantages (Cawood et al, 1988:126-128; White, 1983:131-133; Heywood, 1989:85, 86, 244; Black, 1991:21, 22; Mathews, 1985:22 - 27, 54 - 58, 113 - 134; Courts & McInerney, 1993:3 - 9). In the following discussion the main disadvantages identified by these various authors will be portrayed.

• The first disadvantage is the over-emphasis on facts and on intellectual abilities of learners (Cawood et al, 1988:126; Mathews, 1985:54). In most cases, the questions in standard examinations concentrate on knowledge that can be tested or assessed very easily. White (1983:132) singles out this disadvantage as one of the main ones and he states that a high proportion of the questions only tests the lower cognitive abilities of learners (i.e. knowledge and comprehension) and only the reproduction of facts or the exercising of a simple skill is required.

• As a result of this great emphasis on facts and knowledge, numerous essential life skills and attributes are not promoted or evaluated by the standard conventional examination. Aspects such as social, moral, artistic and physical attributes of learners are not assessed in the examination process nor are these aspects fostered by examinations. The same is the case with regards to skills such as the ability to
manage, to negotiate, to co-operate with others, the skill of versatility, the ability to generate new initiatives, to develop a pride in quality workmanship, the simulation and nurturing of a sense of conscientiousness; effective time-keeping and -management, and so forth (Mathews, 1985:22, 27).

• Cawood et al argue that the choices and preferences of the examiners reduce or limit the syllabi to what they think or feel is important for learners to know (1988:126). In fact, according to Mathews (1985:58) most standard conventional examination questions merely require the recall of information.

• A further disadvantage is the fact that preparations for examinations cause teaching to become nothing more than a mere coaching of learners, the drilling of facts and a process of memorising. This in turn causes a lack of interest in the subject by learners, as well as an absence of insight into the work (Cawood et al, 1988:127).

• Because of this “drilling of facts” learners learn to use wrong and ineffective study techniques. Cawood et al (1988:127) state that such incorrect study methods give rise to the problem that the average learner almost never learn to become an independent thinker or to gain an insight into the work that they have to study.

• Another disadvantage lies in the fact that examinations frequently tend to become an end in itself and learners/pupils are merely trained to pass the examination, but are not being prepared for the practical work situation outside. In many instances, learners regard examinations merely as a means of achieving a certain diploma or degree. They are therefore not concerned about the mastery of knowledge and the acquisition of relevant skills, but rather about passing the examination (Cawood et al, 1988:127, 128). Mathews also mentions this aspect and he feels that examinations do not in fact stimulate or facilitate the acquisition of particular skills. It rather encourages learners only to aim for the obtaining of a particular certificate or diploma or degree, irrespective of whether the skills that are required are obtained or not (1985:24).
• Heywood (1989:244) sees as a disadvantage the fact that an examination is only a single measure of a learner's performance. Such a single measure can rarely give a fair and accurate account of the skills and knowledge that a learner has obtained in a certain subject. A judgement based on such a single measure is therefore grossly unfair.

• A further disadvantage identified by Heywood (1989:244) is the tight time-constraints imposed on learners. He argues that these time constraints can cause learners to perform poorly and because of that they might fail the examination.

• A very serious disadvantage of the standard conventional examination emphasized by a number of authors is the extensive tension and stress caused by examinations (Mathews, 1985:23; White, 1983:133; Heywood, 1989:85, 86). Mathews alleges that this stress and anxiety can lead to frustration, negative personality formations, misbehaviour and even health problems. White regards this disadvantage as one of the main disadvantages of the standard conventional examination. He states that examinations frequently induce high levels of anxiety which can produce abnormal conditions and potentially invalid results. Heywood (1989:85) especially stresses the negative effects of examination stress on learners. He states that it is a well-known fact that tests and examinations cause a great amount of anxiety and stress in learners in the run-up to the dates on which it is to be written. He states that a state of over-anxiety can cause some grievous disturbances in learners which may in turn negatively affect a learner's performance in such a test or examination. Heywood quotes the findings of Malleson (a physician at the London University College) who found that at least 10% of all learners who have to write final examinations, are suffering from strain and anxiety as a result of the looming examinations (1989:86). Malleson's research also came up with the following types of examination-related strain:

  - "Pre-examination strain": This is the anxiety experienced by a number of learners during the preparation (study) phase.
- "Examination panic": This is the anxiety experienced while writing the paper and some learners even get a "black-out" and leave the examination room.
- "True examination phobia": This is an almost irrational fear of examinations that some learners (although a minority) experience (Heywood, 1989:85, 86).

- Black (1991: 21) sees as a further disadvantage the fact that written forms of evaluation cannot effectively serve as a "surrogate" or substitute for practical or real life performance. This view is also echoed by Mathews who says that written examinations as a form of evaluation is in fact not a true reflection of real life, in other words, learners will not at a place of employment be required to display their skills in terms of a particular subject in the same way as required in a written examination (Mathews, 1985:113).

- Another disadvantage identified by Black (1991 : 22) is the fact that summative evaluation cannot effectively include tasks that will clearly convey and illustrate what learners have really learned about a particular subject. It is also almost impossible to exemplify criteria set in examination papers by single definitive tasks as it normally the case in written examinations. Rather a number of tasks should be used to exemplify a particular criterion and this can best be done in formative evaluation (Black, 1991:22). Mathews in fact feels that there is a lot of uncertainty about whether examination questions really adequately give room for learners to display specific skills that the subject might require from them and which form part of the subject practice (Mathews, 1985:57).

- Examinations also creates the problems of "misfit" and "waste" (Mathews,1985:23). That is incidents where learners who would have succeeded in the workplace, could not pass the (final) examination while the ones who do pass excellently, cannot perform the jobs that the certificates obtained, warrant them to.
• Motivation of learners based on examination stress or the compelling influence of examinations, is in fact very negative because it in fact limits the aims of formal education because learners do not effectively learn to generate their own goals and aims internally (intrinsic motivation) but rather rely on extrinsic factors (Mathews, 1985:24).

• The problem of legible handwriting. Learners are on many occasions penalised for illegible handwriting, and on the other hand it causes problems for examiners themselves (Mathews, 1985:114).

• Another serious disadvantage of the standard conventional examination is the problem of bias. This is quite an enormous shortcoming and it manifests itself in manifold ways:

  - The first of these biases in examinations is social. Mathews argues that children from working class parents normally cannot or do not always take advantage of the “competitive examination system” as do the children of socially more advanced parents (1985:121). The reason for this is the fact that examinations are primarily cognitive and intellectually based and because parents from the middle social class are mostly employed in work positions that requires such intellectual and cognitive skills, they are more in touch with such principles and can support their children more effectively. In the case of the manual-worker parent, the opposite is the case (1985:121). Mathews argues that is only possible to prevent such social bias in homogeneous cultures.

  - The second of these biases is gender bias. In the mid 19th century, females were excluded from the examination processes and although they gained access to it at a later stage, for quite same time a “make character” dominated the nature and contents of examination questions and educational curricula (Mathews, 1985:128). There was also the debate on whether girls should write the same examinations than boys and whether they should be confronted with the same criteria and standards (Mathews, 1985:128,129).
A third and probably most problematic bias is the cultural/racial bias in examinations. According to Mathews (1985:133) the Rampton Report that was published in 1981 in Britain, found that a clear euro-centric bias existed in many examination questions that were to be written by "non Europeans" in Britain. This is most probably also the case in South Africa. The question arises to what extent this Euro-centrism has been erased and whether it will ever be possible to erase it from written examinations.

The fourth bias which is closely linked to the third is the linguistic bias. Mathews (1985:133) states that research in Britain has shown that learners who receive their instruction in a second language (i.e. Asians in English) normally perform on average poorer than learners who receive their instruction in their first language. The same dilemma exists in South Africa and the problem is even more severe because it is the majority of learners on tertiary level that receive their instruction in a second language.

In summary it would appear as if the three most serious disadvantages or shortcomings of the standard conventional examination are:

- the over-emphasis on factual knowledge reproduction;
- the high levels of tension and stress that learners experience and the subsequent health and emotional disturbances;
- the presence of various forms of bias.

These three shortcomings are indeed regarded as very serious in this research because it calls the reliability, validity and suitability of the standard conventional seriously into question. These disadvantages would probably be even more severe in as far as the distance learner is concerned because of the fact that learners is not in close contact with the lecturer who could have supplied the necessary guidance and support.
Before a final judgement on the acceptability of standard conventional examinations be issued, the validity and reliability thereof will first be undertaken.

3.5.1.3 Validity of the standard conventional examination

When evaluating the standard conventional examination against the criterion of validity, it becomes clear that especially the three main disadvantages identified in section 3.5.1.2 pose serious threats.

- The first of these threats is also identified by Cawood et al (1988:129, 130) and that is the fact that it is highly unlikely (even impossible) to cover the complete syllabus of a particular subject in one examination paper. The great emphasis on factual recall and on contents that can easily be measured by means of the written examination is already an indication that certain subject contents were excluded from the examination paper. If this is the case, then it means that any argument maintaining that examinations measure the full content of a syllabus, is false. This in turn threatens the validity of examinations, because it would mean in simple terms that examinations do not in fact measure what is claimed that they are intended to measure.

- Second, Cawood et al state that changes in the condition of a learner (this includes physical, mental, social, and so forth.) can also pose a threat to the validity and reliability of examinations as a measurement instrument. If one takes into account that at least 10% of all learners suffer from examination stress, then it means that at least in 10% of cases there is no true reflection of learners' real potential. If one adds to this other possible disturbances in learners' emotional lives (for example if a learner suffered the loss of a beloved (parent, brother, wife and so forth.) very close to the start of the examination. On the day that such a learner has to write a particular examination paper, that learner will not actually be able to respond cognitively to the questions in accordance with what is intended in the paper. Learners might even respond more from an affective mode and in such an instance, the examination paper is not in fact measurement what it is intended to measure.
• The aspect of bias in written examination papers pose a further serious threat to the validity of such an examination. Any kind of bias in an examination paper makes such a paper invalid because it means that there was no consistency with regards to what was measured in terms of different groups of learners.

• Another area of criticism against written examinations mentioned by Cawood et al, is the way in which the questions are formulated. If the questions in a particular paper are formulated in such a way that learners cannot fully grasp what is required from them, then that particular question(s) will not in fact measure the actual aspect it was intended to measure. Such a "misunderstanding" on the side of learners cannot necessarily be blamed on a low IQ or insufficient preparation, but on the format of the question(s). In such an instance there will again be a threat to the validity of that examination.

These threats to the validity and reliability of conventional written examinations, are not the only ones, but indeed some of the main ones. In this research there is total agreement with the abovementioned criticism against conventional standard examinations because these are in fact real threats that were mentioned. These threats are probably also the reason why continuous attempts are made to develop more valid and effective (written) measurement instruments.

This criticism, coupled with the numerous disadvantages as mentioned above, brings one to the conclusion that the conventional standard examination is not in fact the best or most effective measurement technique. However this method of measurement is still in regular use, even at Technikon Southern Africa. In section 3.5.1.5 a comparison of its applicability to the TSA, against that of the other measurement techniques to be discussed here, will be supplied.

3.5.1.4 Reliability of the standard conventional examination

The three main disadvantages of the standard conventional examination also pose some
serious threats to the reliability of this instrument.

- No consistency would be possible if the complete syllabus is not evaluated each time because that will imply that different aspects of the syllabus will be evaluated in different examinations.
- If emotional conditions (as a result of stress or anxiety) always have an influence on the performance of learners, then no consistency will ever be possible.
- If there is any form of bias present in the examination paper, it will always have one or other effect on some of the candidates writing the examination. However, this effect will not be consistent because it might differ from occasion to occasion.

There are also some other threats identified by Cawood et al (1988:129). These are:

- The element of subjectivity which is almost at all times present during the evaluation or marking of examination scripts. This contention is based on the fact that research has proved that different examiners will award different marks for the same paper, even if they mark with the assistance of a carefully compiled memorandum. The reason for this inconsistency, according to Cawood et al. (1988:129) is the fact that individual examiners emphasize different aspects of a learner’s answer higher than others and his mark-allocation is then based upon that particular emphasis. (An example of such aspects is: content, originality, clear expression, own interpretation, and so forth.). The threat that this aspect of subjectivity poses seems to fall more on the level of reliability, because in this case the implication is that examinations do not produce the same results on a consistent basis.

- Another threat to the reliability of examinations, is the aspect of different marking scales (Cawood et al, 1988:120). Research has found that there is a positive correlation between the achievements of learners and the particular marking scale used. If different scales are in fact used to evaluate the work of the same learners on a continuous basis, the same (or similar) results will not be achieved. The
presence as well as the use of these different marking scales pose a definite threat to the reliability of examination marks.

3.5.1.5 Suitability of the standard conventional examination for TSA

In as far as the aspect of distance education is concerned, the standard conventional examination paper does not fully accommodate the distance learner. The reason for this assertion is the fact that the average distance learner will probably "never" be as properly prepared for the examination as an average residential learner. Such a distance learner does not have so much opportunity to consult the lecturer about the finer aspects of such an examination.

• First, cognisance must be taken of the complex nature of the conventional paper (the different types of questions; different mark allocations, etc.). Although such instruments as telephones, etc. are available, the financial implications for the learner living very far from the technikon premises, would deter such a learner from using it. The standard conventional examination paper does not make room for such a disadvantage on the distance learner's side. Written guidance from the lecturer is also not enough for there might even still be certain things that the learner does not understand properly.

• Second, in as far as the aspect of self-study is concerned, the standard conventional examination paper is set on the presumption that the learner understands everything in exactly the same way as the lecturer intended him/her to. Steyn (1989: 451) contests this presumption and argues that in many cases a learner's interpretation of facts and information can be in conflict with that of the lecturer. Especially the objective-type questions which propagate only one correct answer, would put the distance self-study learner at a disadvantage in such instances. It is possible to argue that the lecturer would have had the opportunity to straighten out such "differing" interpretations, but the real life situations might prove different.
• Third, in as far as the practical nature of Technikon Southern Africa is concerned, the standard conventional examination paper is also lacking. The nature of the standard conventional examination is normally very theoretical, because the emphasis is mainly on the recall of factual content (as mentioned in section 3.5.1.2). Such a theoretical emphasis is in fact a misguided one and neglects the practical inclination of Technikon Southern Africa.

• In the fourth instance, the aspect of the mastery of knowledge which would imply the ability to apply that knowledge in new situations, as well as the mastery of certain skills, can not be measured properly and effectively by means of the standard conventional examination. Despite the presence of open-ended questions in a standard conventional examination paper, the scope for questions that would test such mastery, is far too limited to realise that aim. The conclusion reached here is that the conventional question paper is not a suitable instrument to measure the extent of mastery of knowledge and skills.

• In the fifth instance, the second/third language dimension is especially a difficult one. In this respect, all written forms of measurement would place the second/third language learner at a disadvantage. As mentioned in section 3.5.1.2, it was proven that a learner responds optimally in their own mother-tongue. This is one factor that will challenge the suitability of all written measurement instruments.

• Finally, the dimension of multiculturalism poses a further high challenge to measurement instruments. Existing measurement instruments, in accordance with existing methods of education and training, originate from one particular cultural perspective in South Africa, namely a Western orientated one. The standard conventional examination paper is one of the products of this cultural orientation and to maintain that it is culturally "blind" would be factually incorrect. In fact, as mentioned in section 3.5.1.2, research in Britain has proven the fact that there is a definite cultural or racial bias in existing examination papers (Mathews, 1985:133). As a result of this cultural/racial bias certain groups of learners are at an unfair advantage at the expense of others.
To summarize this scrutiny of the standard conventional examination, the conclusion is reached in this research that the standard conventional examination is not the most effective measurement instrument for Technikon Southern Africa. This statement does not intend to deny the value of the standard conventional examination in other areas, but generally speaking it would appear from the above scrutiny that this method of measurement is not the most adequate for Technikon Southern Africa. In view of this conclusion other conventional measurement instruments will also be examined with the intention of determining their adequacy.

3.5.2 Scrutiny of objective questions

3.5.2.1 Advantages of objective questions

A number of advantages of the objective question were identified by various authors.

- About one advantage which can probably be regarded as its main advantage, there is much unanimity among these authors, and that is the reliability of objective questions in terms of scoring (Mathews, 1985:105; White, 1989:78; Habeshaw et al, 1993:35, 36)

- Another major advantage related to its reliability, is its relative objectivity in the marking process (White, 1989:78, 79). This is an area in which objective questions manage to overcome to a reasonable extent the biggest disadvantage of evaluation in education, and that is the problem of subjectivity in the marking/scoring process. Because of clear right or wrong answers the objective question prevents the examiner to make decisions based on subjective judgement.

- A third major advantage of the objective question is its cost-effectiveness, especially in cases where large numbers of learners are to be evaluated (Noah & Eckstein, 1990:86; White, 1989:79; Mathews, 1985:105). Noah & Eckstein (1990:86) for example state that the multiple-choice format of examination became
popular in the US because of its “economic” advantages. It proved to be very useful in the examination of large numbers of pupils because it can be machine scored and data (item) banks can be developed for every subject from which examination questions can be easily formulated.

- In addition to these main advantages some authors mention a few more advantages of the objective question. White (1989:78, 79), Habeshaw et al (1993:35, 36), Mathews (1985:105), Noah & Eckstein (1990:86) mention the following advantages of multiple choice questions:
  
  - The tests can be easily administered.
  
  - The marking process can be done very quickly. These questions are also very easily marked.
  
  - Because of the easy and quick marking process, much bigger parts of a syllabus can be tested at the same time and in turn it helps the educator to have a better record of a learner’s overall achievement in a course.
  
  - These questions also help to focus the attention of learners on some basic course contents which might have been overlooked otherwise; in other words, it helps to establish a more comprehensive evaluation of a syllabus or course.
  
  - It also benefits those learners who have difficulty in a lot of writing (such as essays) and allows them to portray their general knowledge.

In conclusion, it would appear as if it is especially the three main advantages of the objective question that make it such a popular measurement instrument. However, a number of disadvantages have also been identified by various authors as will be illustrated in the next section.
3.5.2.2 Disadvantages of objective questions

Mathews (1985:104, 105) identifies the following major disadvantages of the objective-type questions:

- First he feels that it creates a forced convergent thinking because this type of question does not allow the learner to differ (diverge) from the fixed response required by the examiner. These types of questions in other words force learners to conform. Another author that highlights this shortcoming of objective questions is Steyn (1989:451). Steyn argues that this type of question presupposes a common understanding or interpretation of facts and information by both lecturer and learner. Steyn contests this presumption and argues that it is possible that a learner’s understanding or interpretation of facts might even be in conflict with that of the lecturer. Such conflicting interpretations cannot automatically be regarded as wrong, because a person’s interpretation of facts can be influenced by their cultural background.

- A second disadvantage identified by Mathews (1985:105) is the fact that according to him, the only real objectivity of these questions is located in the consistent and rigid scoring. He feels that they are in essence also subjective, because the questions are formulated by persons and the correctness of answers is also determined by persons which causes an inevitable level of subjectivity.

White (1989:79) points out a number of other disadvantages of the objective-type question, namely:

- He states that objective questions require very high costs to develop it professionally. If this is the case then it would mean that objective questions are not that cost-effective when it comes to the evaluation of smaller groups of learners.
- White states that there are some doubts about the validity of the tests because of the narrow definition of writing (as editing). The question arises whether this
editing-based form of writing can really give a true reflection of learners' writing and other skills. Black (1991: 23) in this regard reasons that performance cannot effectively be evaluated by means of the fixed-response format of objective questions. He feels that such questions make it impossible for examiners to understand learners' reasoning when answering the questions. He argues that those questions contain a very low level of validity because they do not reflect important educational aims, for example, the ability to substantiate a particular argument (1991: 23).

- Another major shortcoming of the objective question, similar such as in the case of the standard conventional examination (see section 3.5.1.2), is the existence of a clear bias in those examination papers in favour of white learners and to the disadvantage of learners of other races (White, 1989:79). As mentioned before, this aspect poses serious threats to the validity, reliability and suitability of this examination format for use at TSA.

A final shortcoming of the objective question that also correlates with one of those mentioned under the standard conventional examination, is the fact that with a few exceptions (i.e. assertion - reason items), these objective - type questions are normally only directed at the evaluation of factual knowledge (Habeshaw et al, 1993:35). The consequence thereof is the fact that only learners' lower cognitive abilities (for example, retention and comprehension) are evaluated, while the higher cognitive abilities (for example, the ability to apply knowledge to new circumstances; to analyse, evaluate, synthesize; give and motivate own interpretations, and so forth) are neglected.

With this summary of the major disadvantages of the objective question, it becomes clear that despite some definite advantages, the objective question still suffers from some major shortcomings that make its extensive use at TSA questionable.
3.5.2.3 Validity of objective questions

Some of the disadvantages mentioned in the previous section holds some serious threats to both the validity and reliability of the objective-type question. These are the following:

- The fact that the narrow definition of writing (as editing) creates doubts about the validity of these questions. To put it in clearer terms, the doubts are based on the uncertainty whether the little amount of writing required from the learners, can give an accurate reflection of the skills learners have mastered and of the objectives that are intended to be evaluated by means of those questions.
- The fact that the tests display a clear bias towards white learners and against learners of other races also causes some doubts with regard to the validity of these questions, because the implication is that these questions do not measure the same aspects in different (cultural/racial) groups of learners because the different groups of learners are differently oriented or inclined towards those questions.

3.5.2.4 Reliability of objective questions

In general terms the objective question can be regarded as fairly reliable (as mentioned in section 3.5.2.1). Some of the reasons for this are probably the following:

- Because of its consistency in results;
- Its relative objectivity in the marking/scoring process.

However, if one considers the threats to the validity of objective-type questions, even its reliability cannot make up for the serious shortcomings.

3.5.2.5 Suitability of objective questions for TSA

In terms of the aspects of distance education and self-study, some serious problems
might arise in the extensive use of the objective question. The reasons for this assertion are first, the fact that learners mostly have to rely on their own interpretation of course material. Such own interpretations might differ (remarkably) from that of the lecturer. Because of the distance between them, the lecturer will not be able to notice such a different interpretation nor be in a position to rectify it in due time. In most cases the first time the lecturer becomes aware of this misunderstanding is when he marks the learner’s examination scripts. Because of the rigidity in right or wrong answers, the learner might fail the examination (not because they did not study enough, but because of misunderstandings).

Concerning the aspect of the practical nature of TSA, the objective question might also run into some problems. The main reason for this is the fact that it is very difficult or complex to construct questions that will effectively force learners to apply their knowledge or that will enable them to demonstrate the practical nature of any subject. In fact, the construction of any objective question requires a lot of training and expertise and even much more in a case where some practical work is to be evaluated. But even in a case where such questions are professionally set, the lecturer will never know exactly how the learner reasoned and how they arrived at a specific answer because of the editing format of writing in objective questions.

Most of the comments with regards to the practical nature of TSA will also apply in the case of the aims of mastery of subject contents. Especially the demonstration of skills mastery will be very a difficult, if not impossible task, by means of the objective question as far as very technical subjects are concerned.

With regards to the aspect of second and third language speakers, the problem is twofold. First in the case of languages, Black (1991:23) questions the possibility of measuring language competency of any learner by means of the objective question. Second, even with regards to other subjects, the fact that there is only one correct answer puts the second or third language user at a tremendous disadvantage. It might for instance be the case that learners select the wrong answer, not because they did not
study hard enough, but because they did not understand the question or the responses correctly because of a limited command of the language in which the paper is set.

Finally, with regards to the aspect of multiculturalism, based on what White (1989:79) and Mathews (1985:133) say, the objective question is totally unacceptable to TSA because then a certain group of learners will be advantaged by this type of question, while others will be disadvantaged.

In conclusion, it would therefore appear when all the relevant factors are considered, that the objective-type question is not in fact suitable for use at TSA.

3.5.3 Scrutiny of subjective questions (short questions)

3.5.3.1 Advantages of subjective questions

The shorter subjective questions lie somewhere between the objective question on the one hand and the essay-type question on the other. Many of the advantages that will be mentioned during the scrutiny of the essay-type question (see section 3.5.4.1) will also be applicable here, although to a much more limited extent. Only a short summary of what is regarded as some of the main advantages of the subjective question will therefore be supplied.

- In contrast to the objective questions, this type of question allows room for a learner’s own reasoning and thinking (although to a limited extent).
- A bigger section of the syllabus can be covered than in the case of the essay-type question.
- The marking process will also be a bit easier than in the case of the essay-type question.
- Rowntree (1990:311) states it categorically that own-answer questions are the most suitable type of measurement to evaluate learners’ higher mental abilities.
3.5.3.2 Disadvantages of subjective questions

The same comments at section 3.5.3.1 are also applicable here with regards to the essay-type question. The following is also merely a brief summary of the main disadvantages of the subjective questions:

- The problem of subjectivity although less than in the case of essay-type questions.
- Not enough room is made for learners' own interpretations and views.
- The required answers are still too prescriptive although they leave some room for own interpretations.

3.5.3.3 Validity of subjective questions

Some of the same threats as in the case of the standard conventional examination will also apply in this case, although to a lesser extent. These threats are:

- The subjective nature of the marking process.
- The impossibility of covering the complete syllabus in one paper.
- The existence of different marking scales.
- Changes in learners' physical or mental conditions.
- The formulation of questions.

3.5.3.4 Reliability of subjective questions

The biggest threat to reliability is of course the subjective nature of the marking process.
The other factors mentioned under the standard conventional examination will also apply in this case. These factors are:

- Different marking scales

- Different emotional/health states.

3.5.3.5 Suitability of subjective questions for TSA

Because the short (subjective) question is in essence a component of the standard conventional examination, the same comments made with regards to the suitability of the standard conventional examination for TSA will also be applicable in this instance (see section 3.5.1.5).

3.5.4 Scrutiny of the essay-type question

3.5.4.1 Advantages of the essay-type question


Duminy & Söhnge (1987:118) list the following advantages:

- Essay-type questions are easy to formulate.
- Learners don’t have difficulty in answering it.
- It cannot easily be replaced as a means of practising and evaluating certain higher mental qualities.
- It evaluates achievement on the interpretation and application levels.
- It helps to develop the following skills in learners:
  - to make summaries of study material;
to distinguish between main and minor issues;
- to discover the relationship between facts;
- to search for ways of applying their knowledge.

Frith & Macintosh (1988:78, 79) also underscore those advantages of Duminy & Söhinge but add a few more, namely:

- The Essay-type question is an effective way to measure a learner’s power of expression in a situation where they are not limited by a set of specific requirements.
- This in turn enables learners to develop own ideas and respond to the questions in the particular way they chooses to.

Another author that mentions something about the advantages of the essay-type question, is Bergman (1981:127). Bergman emphasizes the fact that essay-type questions promote and facilitate the formation and expression of own ideas, and regards this aspect as the greatest advantage of essay-type questions. He further argues that this type of question leads to a more useful and rewarding method of study for learners, because learners in such instances will rather learn ideas, concepts and the general flow of events, than merely a mass of isolated facts.

Habeshaw et al (1993:18) acknowledge some specific advantages of the different types of essay questions that they distinguish. With regard to the open-ended essay question, they mention the following advantages:

- The amount of freedom that this type of question grants learners;
- The opportunity it grants especially the more excellent learner to practice and portray their skills.

With regard to the role play essay they mention the following advantages:
It helps learners to note the relevance of a particular task which in turn might stimulate the learner's personal interest into that task.

Learners' writing tends to become very fluent and natural.

It can positively influence a learner's approach towards the answering of a particular question (Habeshaw et al, 1993: 21).

In terms of the note-form essay they mention the following advantages:

They force learners to concentrate on the essential (main) information.

They are easier and quicker to mark than the other types (Habeshaw et al, 1993:29).

In summary it would appear that the main advantage of the essay-type question is the fact that it allows room for learners' own interpretations of facts and to express their own points of view. It also seemingly allow for the stimulation and evaluation of higher order mental skills such as reasoning, comparing, analysing, evaluation, and so forth. However, not all learners are equally proficient in using the written word to display their higher order skills. There are also a number of other advantages of the essay-type question which will be discussed in the next section.

3.5.4.2 Disadvantages of the essay-type question

It would appear that the greatest disadvantage of the essay-type question, is the aspect of subjectivity (Duminy & Söhne, 1987:118; Frith & Macintosh, 1988:79; Bergman 1981:128). This aspect seems to be much more severe with regards to the essay type question than in the case of the standard conventional examinations, because of the fact that there can be no precise memorandum drawn up to serve as criteria due to the "free" nature of learners' answers.
In addition to this problem of subjectivity, a number of other disadvantages have also been highlighted by different authors. Duminy & Söhinge (1987:118) mention the following:

- Too much writing and too little thinking is done.

- The questions are sometimes indefinite and vaguely worded.

- The marking of such questions takes much longer than in the case of objective questions.

Frith & Macintosh (1988:79) add the following aspects to the list:

- Essay-type questions cover a relatively small area of the subject being studied.

- A single essay does not provide an acceptable measurement of the various skills that it are supposed to measure.

Habeshaw et al (1993: 65,66) mention another disadvantage, that is very prevalent especially in tertiary education, namely the fact that learners tend to “spot” (guess and concentrate on) certain questions for the examination by means of previous examination papers. Learners who are fortunate enough to “spot” correctly, can achieve high marks for the examination. Such a learner is not necessarily a good learner and in many such occasions the particular learner has no insight into the study material. Such a learner will pass the examination, but will not in reality dispose over the minimum requirements to deserve that pass mark.

Another similar disadvantage to that of standard conventional examinations, is the aspect of high levels of tension and stress in learners that normally precede the writing of particular examination papers. The same amount of stress is also present as in the case of conventional examinations.
Cox (1994:90) states that questions that are easy to set (essay type) are normally very difficult to mark, because the process of designing a memorandum and specifically indicating what facts are required or will receive marks, is extremely difficult. Cox (1994:93) states that the marking of essay-type question is therefore a very "arduous" process and very subjective.

Even in terms of the different types of essay-type questions a number of disadvantages can be identified. Habeshaw et al (1993:29) for example identify the following disadvantages of the note-form essay:

- It can unfairly benefit poorly prepared learners because they can merely write anything concerning that topic and still gain marks.

About the open-ended question Habeshaw et al (1993:18) identified the following disadvantages:

- Because it is open-ended, even the unprepared learner might pass because of a 'cobbling' together of disconnected facts.
- Learners tend to learn answers by heart and merely reproduce it in the examination without really understanding its full meaning.
- Because of the normally extensive scope they have to prepare, some learners might panic and even 'stall.'
- This type of question tends to lean towards a great degree of generality.
- There is no absolute clarity on what will count as an acceptable answer.

The list of disadvantages is not exhaustive and a lot more can be mentioned. It is however already clear from the ones mentioned, that also the essay-type questions have some serious shortcomings which hampers its effectiveness as a measurement instrument. The validity of the essay-type question will be examined next.
3.5.4.3 Validity of the essay-type question

In as far as the question of validity is concerned, the essay-type questions creates more room for broader portrayals of skills and objectives set by examiners than objective questions. It might be a bit more likely that the intended objectives to be measured by means of the essay-type question can in fact be measured, provided that the objectives are very clearly and precisely formulated. However some aspects regarded as threats to the validity and reliability of standard conventional examinations (see section 3.5.1.3 and 3.5.1.4), can also be levelled against the essay-type question. These aspects are:

- The subjective nature of the marking process.
- The impossibility of covering the complete syllabus in one paper.
- The existence of different marking scales.
- Changes in learners’ physical, emotional or mental conditions.
- The formulation of questions.

It would therefore appear that the validity of essay-type question is but a precarious matter. With this conclusion, the reliability of essay-type questions will be examined next.

3.5.4.4 Reliability of the essay-type question

The main threat to the reliability of essay-type questions is the aspect of subjectivity. There is no guarantee that another examiner will come up with the same marks as the previous one. The same examiner’s marks even differ on consecutive marking incidences. It appears therefore that in terms of reliability, the essay-type question have some major shortcomings.
3.5.4.5 Suitability of the essay-type question for TSA

In as far as the first criterion is concerned, the essay-type question might also pose a problem in terms of proper preparation for the distance learner. However, in contrast to the standard conventional examination, the essay-type question is less complex in nature and therefore easier to grasp. This by no means, is an argument that preparation for the essay-type question might pose no difficulties to the distance learner, but rather that it would be easier to understand independently.

Also in terms of the second criterion, the essay-type question might be easier to implement than the standard conventional examination. Although this type of question might also be set up with the presumption that learners understand and interpret information the way the lecturer intended, it does allow learners to state and motivate their own ideas and views on the information. In fact the essay-type question makes excellent provision for self-study learners to convey their own interpretation and understanding of particular knowledge contents.

In as far as the third criterion is concerned, the essay-type question does narrow the gap between the theoretical and the practical, because by employing higher cognitive skills such as application, analysis, synthesis and evaluation, a move is made towards the practical application of knowledge. However, the fact that it is a matter of "writing about" a certain process makes it in essence still a theoretical measurement instrument. A purely practical oriented measurement instrument would be more concerned with the actual "doing" or "performing" of a particular task.

In as far as the evaluation of mastery of knowledge and skills is concerned, the essay-type question is less bound by limitations as the standard conventional examination in as far as the practical component is concerned. The activation of the higher order skills can possibly illustrate the extent of mastery of certain skills and knowledge. However various subject disciplines would require the "doing" or "performing" dimension in order to accurately determine the extent of mastery of particular skills and knowledge.
Concerning the language aspect, the essay-type question in accordance with the standard conventional examination being written forms of measurement, also places the second or third language learner at a disadvantage. In fact, the problem would tend to be more severe in the case of the essay-type examination because of the fact that a learner would need a good command of a particular language in order to bring across his ideas in an effective way. For a second or third language learner, this might cause some problems.

In the last instance, the cultural origin of this measurement instrument might not cause such an extensive problem as in the case of the standard conventional examination. The reason for this contention is the “open-endedness” of the essay-type question. In for example cases where no specific (culturally bound) answer is expected from learners, it will grant them the freedom to express their own ideas and interpretations.

In conclusion it would appear that the essay-type question is capable of being an improvement on the standard conventional examination. However, as indicated in the case of practical, mastery and language criteria, it can still not be regarded as the best or most effective measurement instrument.

3.5.5 Scrutiny of the structured essay-type question

3.5.5.1 Advantages of the structured essay-type question

Habeshaw et al (1993:23) regard as the main advantage of the structured essay, the fact that it makes the marking process much easier.

Frith & Macintosh (1988:69) mention a number of advantages and characteristics of the structured question, namely:

- These questions provide learners with very definite signposts for the writing of their answers. In other words, it gives particular direction and parameters to learners’ answers.
Second, these questions can also provide certain stimuli to learners which would help to bring their thinking processes into operation.

These questions can be used for a wide variety of purposes, including the measurement of certain skills or the extent to which such skills were mastered.

The structural aspect has also encouraged the organisation and interrelationship of essentially short-answer questions with large sets of knowledge contents.

Mathews (1985:108) also list a number of advantages of the structured question, but these advantages are especially to the benefit of the learners. Here follows some of the main advantages that he identifies:

- The structuring of the questions makes it easy for the learners to know what the requirements of the examiner are.
- It also guides the learners in terms of how much information is required for each heading and what the weight of each heading is (in terms of the marks it counts).
- Because of the information normally presented in the stem of the questions, it makes application type questions more possible instead of mere recall questions.
- It makes it possible for the educator to break complex subject matter into more understandable parts for the learner, thus enabling the weaker learners also to be able to achieve some marks for the question.
- These question are therefore particularly suitable for learner populations with a wide variety of cognitive and academic abilities and skills, because it makes it possible for all types to be able to answer most of the questions.

In addition to these advantages, some of the advantages of the essay-type question might also apply here to a limited extent. The advantages mentioned by these authors are acceptable and can probably withstand some criticism. However this type of question does place limits on learners' freedom to express their own ideas and arguments liberally. It also tends towards the direction of the objective type of questions which presupposes a common understanding of what is right and wrong.
With these comments the actual disadvantages of the structured question will be examined next.

3.5.5.2 Disadvantages of the structured essay-type question

A number of disadvantages with regard to the structured essay-type question can be identified. Here follows a brief summary of those disadvantages:

- Frith & Macintosh (1988:70) mention as a possible disadvantage of structured questions, the fact that they contain fertile ground to become too lengthy and not specifically focused on the essentials.

- The aspect of subjectivity in marking would also apply to the structured question, because although there are certain limitations and barriers included in the questions, it does allow for free expression to a limited extent. It will again be difficult, if not impossible to compile an inflexible memorandum which must be followed to the letter. For this reason, the examiner's subjective inputs in marking cannot be denied.

- Another disadvantage is the fact that the structured question does not allow the same extent of freedom to learners to express themselves liberally. Limitations are set by means of the headings and particular emphasis that the questions require. These limitations also prevent a proper illustration of skills as well as the extent of mastery of those skills, because of the relatively short nature thereof (in comparison with the essay-type question).

Mathews (1985:108) also identifies a few disadvantages of the structured question. Those disadvantages are the following:

- They do not allow so much room for imaginative writing as does the open-ended essay-type question.
- They do not allow for exact or objective marking as does the objective question.
In conclusion it is sufficient to say that the structured question is an improvement on the standard conventional examination because it moves a bit away from the rigidity and presupposition elements of the standard conventional examination (objective questions) to a more freer, less prior determined interpretation of knowledge. However, the limitations, as mentioned, do not make it a very effective instrument for the development and measurement of higher order mental skills. Especially for a learner at tertiary level, the limitations will be too contra-productive.

With this concluding remarks, the validity and reliability of the structured question will be touched on next.

3.5.5.3 Validity of the structured essay-type question

The same aspects mentioned in the case of the essay-type question would also apply to the structured question with only slight variations. For that reason the reader is referred to section 3.5.4.3 in this regard.

3.5.5.4 Reliability of the structured essay-type question

Refer to section 3.5.4.4.

3.5.5.5 Suitability of the structured essay-type question for TSA

Because the structured essay-type question is merely a particular type of essay question, the same comments made in the case of the essay-type question, will also apply here (see section 3.5.4.5).
3.5.6 Scrutiny of the written case study

3.5.6.1 Advantages of the written case study

Because of the fact that in the written case study a variety of question types can be used, only the specific advantages of the case study format will be examined. The following is a brief summary of such advantages:

- Because of the variety of questions that can be used, a bigger part of the syllabus can be covered in one examination.

- It creates opportunities to expose learners to simulations of real life problems experienced in the business world.

- It creates opportunities for learners to apply their theoretical knowledge to realistic types of problems.

- It can also be regarded as an improvement on the standard conventional examination because it represents a move away from the mere factually-based type of question.

3.5.6.2 Disadvantages of the written case study

The main disadvantage of the written case study is the fact that it cannot in effect transcend the limitations of the written medium. The shortcomings that remain can be summarized as follows:

- It is still merely a type of written examination which causes even those application questions to remain theoretical.

- The application of skills remain superficial because it is expressed by means of the written word, while in real practice it has to be physically demonstrated.
3.5.6.3 Validity of the written case study

- The first threat to validity will be the fact that (as mentioned) the examination paper will not in actual fact be able to measure the skills that will be required in real practice and that are probably intended by an examiner.

- Other threats are the same as those mentioned in section 3.5.1.3.

3.5.6.4 Reliability of the written case study

It would appear that the main threats to the reliability of the written case study are the following:

- The problem of subjectivity in the marking process.

- The fact that different types of questions can be used makes it susceptible to inconsistencies.

3.5.6.5 Suitability of the written case study for TSA

In terms of the first two aspects, distance education and self-study, basically the same comments as in the case of the standard conventional examination (section 3.5.1.5) will be applicable. However, with regards to the aspects of the practical nature of TSA and the primary aims of subjects to get learners to a stage of mastery of subject contents and skills, the written case study (and the open book examination) come a bit closer to a realisation of those aims than do the other measurement instruments discussed here. This assertion is based on the fact that the case studies (although simulations) can give a reasonably accurate picture of real life situations in practice and by so doing, it brings learners closer to real practice. It also creates room for a demonstration of skills and the extent to which those skills have been mastered by learners. However, this is still
only to a limited extent because it remains essentially theoretical because of the written and superficial nature thereof.

With regard to the remaining two factors, second and/or third language learners and multi-culturalism, the written case study also suffers from the problem of linguistic and cultural/racial bias because it is still a written form of measurement with a western basis in it.

So in conclusion it would appear that although the written case study succeeds to a limited extent to overcome some of the short-comings of the other written forms of measurement, it can still not be regarded as the most suitable form of measurement at TSA or for Business Management.

3.5.7 Scrutiny of the open-book examination

3.5.7.1 Advantages of the open-book examination

Heywood (1989:243) identifies the following advantages of the open-book examination:

- It helps to reduce examination stress.

- There is no drastic differences in learner achievement noticeable when open book examination are administered.

- It improves and enhances learners' preparation for the examination.

- The open book examination makes it possible that different abilities (higher order) are tested and not memory only.

- It also contributes positively to the level of validity and reliability of the examination itself.
It also leads to a decline in learner cheating.

White (1983:132), Frith & Macintosh (1988:79, 80) and Bergman (1981:82, 83) list an additional number of advantages of the open-book examination, namely:

- It assesses the higher levels of learners' cognitive abilities which are, for example, the ability to:
  - comprehend and gain an insight into study material;
  - assemble data and to discriminate between important and less important information;
  - express their own ideas;
  - to evaluate and synthesize;
  - to arrange or organize data in a logical and coherent manner.

- It reduces the emphasis on factual recall (retention of facts).

- It encourages learners to become familiar with their textbook and tests their ability to locate, select and use information.

- The problem of copying or external assistance will be eradicated completely because the answers required cannot be copied or prepared prior to the examination.

- This examination type is closer to reality and practice, because many work situations require the preparation of reports where access to and the use of reference material is required.
This list is not exhaustive and a number of other advantages can probably be added. However, these lists of advantages provided in this section give a reasonably fair account of the main advantages of the open-book examination.

An examination of the possible disadvantages of the open-book examination will be embarked on next.

3.5.7.2 Disadvantages of the open-book examination

Bergman (1981: 83) maintains that there are a number of disadvantages to the open-book examination, namely:

- The fact that open-book examinations are more difficult to grade than other types of examinations.

- Greater demands are placed on the lecturer's time than in the case of standard conventional examinations, because the lecturer will be required to have an excellent understanding not only of the study material but of all relevant material in order to evaluate learner responses adequately.

- The average grades tend to be much higher than in conventional and essay-type examinations, and this deters lecturers from using it extensively.

- It is also difficult to discriminate between better and average learners, because this will entail an attempt to assign objective grades to subjective performances.

In conclusion it would appear as if the disadvantages mentioned by Bergman appear to be more lecturer-centred than learner-centred. In other words, the main disadvantages of the open-book examination are with regard to educators rather than learners. Other disadvantages attached to the open-book examination are also with regard to the written nature of the examination (see section 3.5.1.2).
At this stage, a scrutiny of the validity and reliability of the open-book examination will be undertaken.

3.5.7.3 Validity of the open-book examination

In as far as validity is concerned, the open-book examination can come very close to measuring the prior set objectives of a particular examination or course. The reason for this conclusion lies in the fact that the main aim or intention with open-book examinations appears not to be the measurement of the retention of subject matter and content, but the extent to which certain skills were mastered by learners. Unfortunately only a limited number of those skills that will be required from learners in a real working environment can be measured by means of the open-book examination because of its written nature.

It is also true that threats such as subjectivity in marking; the physical and mental conditions of learners; different marking scales, and so forth, are also present in the case of the open-book examination, although to a lesser extent than in the case of the other forms of written measurement.

In conclusion it would be fair to reason that the open-book examination can be a relatively valid measurement instrument, provided it is limited to those skills that can be measured by means of the written word.

3.5.7.4 Reliability of the open-book examination

The main threat to the reliability of the open-book examination is the aspect of subjectivity in the marking process. Other threats that might exist will probably be similar to those in the case of the essay-type question (see section 3.5.4.4).
3.5.7.5 Suitability of the open-book examination for TSA

In terms of the first criterion, the open-book examination would facilitate and accommodate the distance and self-study learner more than any of the previous measurement instruments. The reason for this assertion is the fact that the open-book examination has a central component the testing or measurement of learners' ability to handle and work with their study material. So the proficiency (or lack of it) with which the learner goes about with their study material, would be an indication of the extent to which that learner has mastered the particular skill (the skill of working efficiently with material) and would therefore be one the aspects intended to be measured by this examination. Unfortunately, as mentioned in section 3.5.7.2, it is only a limited number of skills that are required in real practice that can be measured by means of the open-book examination.

In as far as the second criterion is concerned, the same would apply as was mentioned in terms of the essay-type question.

In terms of the practical criterion, the open-book examination comes a step closer to practical measurement, because learners must for instance show in practice how they work with their study material. The open-book examination also entails the employment of higher cognitive skills as mentioned in section 3.5.7.1. The open-book examination can in fact to a lesser extent be regarded as a practical examination. This measurement instrument does in fact bring theory and practice closer to each other, however to a very limited extent.

The open-book examination also provides greater room for the measurement of skill and knowledge mastery. It can therefore be regarded as a move towards more suitable measurement instruments that does not only measure factual knowledge retention. However, it seems as if the open-book still has a long way to go in terms of the conditions of TSA.
The language aspect, however, also appears to be a hurdle in the case of the open-book examination. Just as in the case of the essay-type examination, learners without a proper command of the language in which they have to answer the questions, will still find it difficult to formulate and express their own ideas effectively.

The cultural bias can be minimized if the examiner is not culturally bound when marking and evaluating learners' scripts but allows for a variety of cultural-based ways of expression.

In conclusion it would therefore appear that the open-book examinations come fairly close to an effective measurement instrument for Technikon Southern Africa. However, the language barrier and its limitations with regard to its written nature, prevents it from being recommended as an acceptable measurement instrument for TSA.

With this conclusion in mind, the final written measurement instrument will be evaluated next.

3.5.8 Scrutiny of the pre-set question

3.5.8.1 Advantages of the pre-set question

Heywood (1989:244) identifies a number of advantages in the pre-set or seen examinations, namely:

- It helps to reduce anxiety in learners.
- It enhances the motivation of learners.
- It can be used to help develop learners' learning skills.

Habeshaw et al (1993: 65,66) add the following advantages of the pre-set question:
• It eliminates the aspects of spotting and prevents learners to pass merely on the basis of luck.

• Learners' answers are of a higher quality than in unseen examinations.

• It achieves two purposes, namely letting learners write an examination as well as doing their course notes beforehand.

Some other advantages that can be observed in the use of the pre-set question, are the following:

• Learners will have ample time to prepare a good answer. They will have the opportunity to gather all the necessary information and or facts that they would prefer to include in their answers.

• Learners will be given the opportunity to practice and develop higher cognitive skills such as the ability to organise, to structure, to argue logically and to motivate their arguments, to compare, to discriminate, and so forth. Although this same aspect is also true in the case of the essay-type question, the limited time constraints of an examination do not allow a very optimal employment and development of those skills, as in the case of the pre-set question.

Despite these advantages, the pre-set question also contains some disadvantages as will be illustrated in the next section.

3.5.8.2 Disadvantages of the pre-set question

Some of the disadvantages identified in the case of the essay type question (see section 3.5.4.2) will probably also apply in as far as the pre-set question is concerned. Some additional disadvantages can however be pointed out, namely:
• Learners can consult the help of others (more experienced and highly capable people) when preparing the answer for the examination.

• Learners can, after preparing the answer, learn it so by heart that they can run the risk of getting a blank in the course of the examination which might cause them to fail the paper.

• Learners will not make an effort to study and get the rest of the work under control and they might regard it as less important.

Habeshaw et al (1993: 65,66) add the following disadvantages of the pre-set question:

• It can disrupt other courses that do not use the unseen examination.

• It places a lot of pressure on libraries to have enough copies of certain books available.

With this summary of some main disadvantages of the pre-set question, the validity of the pre-set question will be examined next.

3.5.8.3 Validity of the pre-set question

In addition to the threats to validity of the essay-type question (see section 3.5.4.3) the pre-set question is probably subject to another main threat, namely (as mentioned under the disadvantages) the fact that it might not be learners' own work that would be assessed. This poses a threat to both the validity and reliability of the pre-set question:

• In as far as validity is concerned, this will imply that the examination is not measurement what (or in fact “who”) it is intended to measure, because the intention would for example be to measure the extent of mastery of certain skills of particular learners and not the skills of other people who have mastered those skills long ago.
3.5.8.4 Reliability of the pre-set question

In addition to the threats to the reliability of the essay-type question (see section 3.5.4.4) the aspect of learners making use of others’ help, can cause the following additional threat:

- The fact that there will be no consistency at all, because learners would probably employ the help of different people at different times.

In conclusion it would appear as if in terms of both validity and reliability, the pre-set question can face some major threats. With this comment in mind, the suitability of the pre-set question will be examined next.

3.5.8.5 Suitability of the pre-set question for TSA

The same comments issued in the case of the essay-type question (see section 3.5.4.5) would probably also basically be applicable to the pre-set questions, with slight variations in some instances.

In as far as the distance aspect is concerned, the pre-set question will help to minimize the problem of proper preparation, because the learners’ prior knowledge and insight into the question(s) will enable them to lessen the effect of any possible structural problems that might be experienced.

In as far as the self-study aspect is concerned, the pre-set question would in fact be ideal suitable, because the advance preparation of his answer(s) will give him/her the opportunity to clarify and motivate their arguments logically and comprehensibly.

In as far as the practical and mastery criteria are concerned, exactly the same comments in the case of the essay-type question will apply.
The language criterion might appear to be even more problematic than in the case of the essay-type question. The pre-set question might easily lure the second/third language learner into consulting of someone more fluent and proficient in the language in which the question(s) must be answered. This, as indicated in section 3.5.8.3, can pose a threat the validity of the pre-set question.

With regards to the last criteria, the comments issued with regards to the essay-type questions will again be applicable.

In summary it would seem that the pre-set question is capable of having a few additional advantages than both the conventional and the essay-type question. However, the danger indicated with regards to the second/third language learner and the serious problem of outside assistance, limits the effectiveness of this measurement instrument. Added to this is off course the limitations of the written form of examination.

3.5.9 CONCLUSION

Based on the above scrutiny of the various summative measurement instruments in current use, a number of observation are forthcoming, namely:

- It appears that written measurement instruments, despite certain definite advantages, pose some serious difficulties or obstacles for certain groups of learners because of an apparent bias towards certain cultural/race groups.
- In terms of suitability for TSA conditions it seems to appear as if none of the written measurement instruments adequately cater for those conditions which in turn causes unfavourable conditions for the average learner.
- Even in terms of the most basic requirements of any educational measurement instrument, namely validity and reliability, it appears as if there are some serious shortcomings with regard to the written summative measurement instruments discussed in this chapter.
As a result of these observations, the assumption is made by the author of this research that reasonable adequate proof was given to substantiate the allegation made in chapter 1 (section 1.3) about the doubts concerning the continued suitability of the current summative measurement instruments used in Business Management at TSA.

With these few comments this chapter will be concluded and the next chapter will be devoted to an examination of the available alternative measurement instruments and specifically those that might be suitable for use in Business Management at TSA.

3.6 SUMMARY

In this chapter the following aspects were discussed:

- A definition of what is meant by the concept, conventional summative measurement instruments, was given;
- A description of the most common summative conventional measurement instruments was given;
- A global contextualization of the most commonly used summative measurement instruments was supplied;
- A brief exposition of the basic measurement instruments used in Business Management at tertiary institutions in South Africa was provided;
- The measurement instruments used in Business Management at TSA was also illustrated;
- Finally, a basic scrutiny of the most common conventional summative measurement instruments was supplied.

The next chapter will be devoted to a search for and an examination of alternative measurement instruments that might be suitable for use in Business Management at TSA.
CHAPTER 4

ALTERNATIVE EDUCATIONAL MEASUREMENT INSTRUMENTS

4.1 INTRODUCTION

In the previous chapter a description and evaluation of various conventional educational measurement instruments were given. In this chapter an attempt will be made to describe and evaluate some of the most dominant alternative educational measurement instruments that are already being used elsewhere.

This chapter will start off by elaborating a bit more on the definition of alternative measurement instruments as supplied in chapter one as well as a summary of the reasons why a need for alternative instruments developed in educational circles. Thereafter, a list of the various alternative instruments that will be examined in this chapter, will be supplied. Following this list of alternatives, each alternative instrument will then briefly be described in terms of what it entails and how it operates and thereafter each instrument will be evaluated with regard to its advantages, disadvantages, suitability for TSA, suitability for Business Management and its financial implications. Following this an attempt will be made to synthesise the information and findings reached after this process of evaluation. In conclusion, this chapter will end off with a summary of the main aspects discussed in the chapter.

4.2 WHAT IS MEANT BY ALTERNATIVE FORMS OF EVALUATION/MEASUREMENT?

Various terms are used to describe alternative measurement instruments and the following is a list of the most common terms:

- Alternative assessment
- Authentic assessment
- Curriculum-based/Curriculum-embedded (performance) assessment
Dynamic assessment
Portfolio assessment
Performance assessment

For the purpose of this chapter the term, measurement, will be used instead of assessment in order to help focus the attention on that section of the evaluation process that is the main concern of this research.

4.2.1 Alternative measurement

Alternative measurement according to Khattri & Sweet (1996: 2) is a term used to primarily denote measurement instruments different from the traditional and multiple-choice types of measurement.

4.2.2 Authentic measurement

Johnsen (1996: 12) describes authentic measurement as a form of measurement which reflects the actions of learners in realistic situations.

Fisher & King (1995:2,3) define authentic measurement as: “an inclusive term for alternative assessment methods that examine students’ ability to solve problems or perform tasks that closely resemble authentic situations”. They state that authentic measurement focuses on what learners do and how they accomplish certain tasks.

Authentic measurement according to Khattri & Sweet (1996: 2) refers to an approach to measurement which highlights the real world nature of tasks as well as the contexts of those tasks.

Authentic measurement according to Gipps (1994:98), is a form of performance-based measurement which is carried out within an authentic context. Authentic refers to measurement tasks that are performed as part of students’ normal work and not as specific evaluation tasks. Gipps regards portfolios as a good example of authentic
measurement because portfolios include examples of actual student performance which occurred during normal classroom practice. Gipps (1994: 98, 99) further states that authentic measurement can also be specified in terms of a specific aspect, such as: the stimulus; complexity of the task; the locus of control; motivation; spontaneity; resources; conditions; criteria; standards; and consequences.

4.2.3 Curriculum-based /Curriculum-embedded (performance) measurement

The alternative terms, curriculum-based and curriculum-embedded measurement seem to have more or less the same meaning, although there might be some differences in terms of emphasis.

Gearheart & Gearheart (1990: 82, 83) who use the term curriculum-based measurement, emphasise the principle of “alignment” as one of the crucial elements of this type of measurement. This alignment according to them should exist between evaluation, instruction and the curriculum in order to achieve optimal learning by learners.

Gordon & Bonilla-Bowman (1996:36) use the term curriculum-embedded performance measurement. They describe it as a process in which almost no separation is made between teaching, learning, and evaluation. All these processes are regarded as a continuously interactive process in which both instruction and evaluation takes place. Curriculum-embedded performance measurement also requires that learners be involved in real tasks and activities.

4.2.4 Dynamic measurement

The concept, dynamic measurement, is described by Gearheart & Gearheart (1990: 4) as referring to a type of measurement which focuses on the psychological elements involved in the learning process. They also state that dynamic measurement is sometimes called mediated or learning potential measurement.
4.2.5 Portfolios

Portfolios are also commonly regarded as a form of alternative measurement (Kamen, 1996: 859; Johnsen, 1996: 12), but because of the fact that many authors regard it as a type of performance-based measurement (Priestley, 1982: 225; Gipps, 1994:98) it will be discussed under the heading of performance-based measurement.

4.2.6 Performance-based measurement

It appears as if the term, performance-based measurement, includes the main elements of all the various terms which refer to alternative forms of evaluation (Khattri & Sweet, 1996: 2, 3). Authors such as Burry-Stock, Shaw, Laurie & Chissom (1996: 252) and Johnsen (1996: 12), point out the interwoven relationship between the various terms discussed in section 4.2. All agree on the fact that the most common or basic element in all the various terms is the aspect of performance. So it appears as if performance-based measurement is the most suitable term to represent all forms of alternative measurement discussed in this section.

Kamen (1996: 859) regards all these terms as descriptions of the move away from the traditional forms of measurement to more meaningful forms of measurement. Kamen (1996: 860), referring to Worthen (1993a), suggests two common elements in all these various terms, namely:

• all these terms are regarded as alternatives to traditional tests (which include multiple-choice and standardised achievement tests);
• all these terms involve direct measurement of student performance with regards to tasks that reflect actual tasks in real job situations.

4.2.7 Conclusion

An analysis of the various definitions of alternative forms of evaluation/measurement, shows that the most dominant common feature or element in all of them, is the aspect of performance. Because of the central role of performance in the various forms, the researcher prefers to use the term performance-based measurement to include all the
various terms that refer to alternative forms of evaluation. In the rest of this chapter only the term performance-based measurement or performance-based measurement instruments will be used to indicate alternative forms of evaluation. A detailed definition, description and evaluation of performance-based measurement will be provided in section 4.4.

4.3 WHY THE NEED FOR ALTERNATIVE MEASUREMENT?

The need for alternative measurement instruments was already mentioned in chapter 1 (section 1.2 and 1.3). Below is an attempt to orderly list the main reasons that caused a need for alternative educational measurement instruments.


- Various authors mention the growing rejection of multiple-choice and other traditional forms of measurement and a search for more meaningful and substantial forms of measurement. They emphasise the need to evaluate the total process of student performance and the particular skills used to achieve a certain result (Broadfoot, 1986: 2; Hall, 1989: 16; Khattri & Sweet, 1996:1 - 11; Wiley & Haertel, 1996:65 - 67).

- The greater demand for thinking skills in industry and other sectors (Fisher & King, 1995: 6, 7).

- The need for the teaching of essential life and job-related skills to enable learners to meet the challenges of an industrialised world (Broadfoot, 1995:9, 10).

- There are also a few less dominant yet important reasons, such as, for example, the need for optional approaches that will accommodate the various learning styles of
various learners, for example some learners are more theoretically inclined while others are more practical or pragmatic; some are dependent on their senses in the learning process, and so forth (Entwistle, 1988:67).

Khattri & Sweet (1996:3) propose two other main reasons why a strong demand for alternative and specifically, performance-based measurement, developed in recent years. These reasons are:

- The emergence of a constructivist model of learning in the cognitive sciences.
- The demand amongst the business community for learners with proven (assessed) basic skills that will enable them to compete in the modern global economy.

Proponents of reform in educational evaluation in the US, insisted on the following three requirements of evaluation methods:

- A generative view of knowledge should form the basis of evaluation practices.
- Learners should be actively involved in the process and be able to produce or perform something or some task.
- The tasks that are to be performed should be meaningful and not mere easily scorable or easily testable subject contents (Khattri & Sweet, 1996: 4).

Linn & Baker (1996:84, 85) provide the following list of reasons why a move away from multiple-choice testing to performance-based measurement was taken in the US:

A belief or demand that:
- measurement tasks should be authentic;
- measurement methods should correlate with reforms in the curriculum;
- evaluation activities should promote and enhance the instruction process;
- evaluation methods should even contribute towards staff development;
- evaluation activities should be meaningful and intrinsically motivating to learners;
- evaluation practices should be more accountable than the existing ones.
In conclusion, it appears as if there are numerous and probably quite valid reasons why attempts are increasingly made to find alternative forms of measurement. Within the context of Technikon SA (TSA) and specifically Business Management (BM), a similar need exists. With this contextualization of the move towards alternative measurement instruments world-wide, a more focused discussion of performance-based measurement will be undertaken.

4.4 PERFORMANCE-BASED MEASUREMENT

As mentioned in chapter 3 (section 3.4.2) there seems to be a major shift towards the use of performance-based measurement methods in general (including portfolios, profiles and records of achievement in the US as well as in the UK) (Gipps, 1994: 108 - 112; Assiter & Shaw, 1993: 19 - 21; Broadfoot, 1986: 1 - 11; Hall, 1989:14 - 17). Broadfoot (1995: 12 - 18) also describes the phenomenal growth of performance-based measurement in countries such as Australia, Britain and the US.

Khattri & Sweet (1996: 5 - 11) claim that the performance-based measurement movement which gained its momentum in the early 1980's, has made great strides in the US. Performance based-evaluation according to them, is promoted and recognised on national, state, district and school levels. Strong & Sexton (1996: 102) also testify to this effect. They state for example that almost half of all the states in the US were either in the process of implementing or exploring or using performance-based evaluation programmes.

4.4.1 Definition

Gipps (1994: 99) provides the following definition of performance-based measurement which states that:

- "Performance assessment is defined as a systematic attempt to measure a learner's ability to use previously acquired knowledge in solving novel problems or completing specific tasks. In performance assessment, real life or simulated
assessment exercises are used to elicit original responses which are directly observed and rated by a qualified judge.”

Rivera, Kuehne & Banbury (1995: 34) define performance-based measurement as:

- “Performance assessment can be defined as an activity or response that requires the learners to integrate various skills and levels of understanding to solve a problem”

Fisher & King (1995: 4) describe performance-based measurement as a process that requires the performance of actual behaviour or tasks by learners and not merely the answering of questions.

McBeath & Lassen (1992: 283) state that performance tests normally require the demonstration of skill mastery and the application of skills. It can include aspects such as problem-solving, task performance and other activities that reflect real life tasks and skills.

In conclusion it appears that performance-based measurement focuses on a visible demonstration of actual skills that should have been mastered by learners in a particular subject or context.

4.4.2 Different types of performance-based measurement

The following is merely a brief summary of the various performance-based educational measurement instruments as proposed by authors such as Priestley (1982:17 - 43); Gearheart & Gearheart (1990: 81 - 94); Broadfoot (1986: 1 - 11); Khattri & Sweet (1996: 2, 3, 5); Gipps (1994:98) and Hall (1989: 15, 16), and which will be examined in this chapter.

- Actual performance measurement
  - Work-sample tests
- Identification tests
- Supervisor ratings
- Peer ratings
- Self-evaluation

• Simulations
  - Simulated performance tests
  - Simulated identification tests
  - Written simulations
  - Management exercises
    - Role-playing
    - Simulated interviews
    - Fact-finding exercises
    - Case studies
    - Leaderless group discussions
    - Other types

• Observational measurement
  - Checklists
  - Rating scales
  - Anecdotal records

• Oral measurement
  - Oral examinations
  - Interviews
  - Prepared presentations

• Programme requirements
  - Personal records
  - Performance records (profiles, portfolios)
  - Profiles
  - Records of achievement
With this brief illustration of the various performance-based measurement instruments that will be examined in this chapter as background, the first aspect concerning performance-based measurement that will be discussed is some general characteristics thereof.

4.4.3 Characteristics of performance-based measurement

Gipps (1994: 99) states that some of the central elements of performance-based measurement are:
- a direct demonstration of capabilities by examinees;
- a great reliance on observation.

Khattiri & Sweet (1996: 5) mention the following characteristics of the performance measurement process:
- Learners have to structure the measurement task.
- They have to practically apply given information.
- Responses should be well-constructed.
- Learners are also required to explain the process they followed to reach certain answers or conclusions.
- Performance-based measurement is normally rated on the basis of set criteria.

Walklin (1991: 6, 7) lists the following aspects as criteria for performance:
- standards
- behaviour
- conditions
- evidence of ability to perform work (competence)
- the indication of required supportive knowledge and understanding
Walklin regards standards and competence as the two most crucial performance criteria. He defines standards as:

- "a specification by which the qualities required of something may be tested or compared",

and competence as:

- "the demonstration of knowledge and attitudes required to perform a given task or act".

In the context of a particular job context he defines competence as:

- "the ability to perform occupational or work-related activities and to demonstrate underpinning knowledge, skills, understanding and personal effectiveness to standards required" (1991: 6).

Linn & Baker (1996:87 - 89) distinguish the following characteristics of performance-based measurement:

- The tasks in performance-based measurement are normally open-ended and not fixed-choice.
- Performance-based measurement normally involves the use of higher cognitive skills such as problem-solving, reasoning and so forth.
- Performance-based measurement tasks usually require longer periods of time to complete and can range from a few minutes to a few months.
- Some performance-based measurement activities require the participation of groups of learners.
- Performance measurement allows for greater room for choice by learners.
- The scoring process relies heavily on human judgement.

McBeath & Lassen (1992: 286 - 296) state that there are four essential elements common to all forms of performance measurement. These elements are:

- Observable performance: They state that the performances of learners are active, participatory and observable, not mere descriptions (1992:286).
- Domain of performance: They state that performance testing can focus on any one or all three of the main domains, cognitive, affective or psychomotor. The objectives of the test will determine what domain will be focused on (1992: 289).
• Standards for performance: These standards include some or all of the following aspects:
  - The accuracy level with regard to the product.
  - The accuracy level with regard to the process.
  - The requirement of proper sequencing of actions or steps.
  - The need for particular time limits (1992: 290).

• Conditions for the performance test: They state that the conditions or setting where a performance test is conducted, should reflect similar aspects as in a real situation. These conditions include aspects such as the following:
  - physical environment;
  - equipment, materials or instruments;
  - required guidelines;
  - other people where necessary;
  - the purpose of the test and the required performance;

Rivera et al (1995:34) give a summary of some characteristics. They include:
• the facet of application of skills;
• tasks that have to be performed in a real or simulated setting;
• specific criteria for the evaluation of the performance or product.
They also mention the fact that performance-based measurement essentially involves "active hands-on" involvement of learners (1995: 40).

In conclusion it would appear that these numerous lists of characteristics all focus on the central element of visible demonstrations of skills acquired by the learners involved. The emphasis also seems to fall on the aspect of meaningful tasks in which the higher cognitive skills of learners (for example, logical reasoning, problem-solving, substantiation of arguments, critical thinking and so forth) will be promoted and developed.

With this conclusion in mind the aspect of validity of performance-based measurement will be examined next.
4.4.4 Validity of performance-based measurement

Gipps (1994:100) argues that performance-based measurement has a high level of "face validity" because it assesses practical tasks that can be seen. However, she points out that critics do not regard face validity as a sufficient indicator of validity (1994: 102). She therefore distinguishes two further forms of validity, namely construct and consequential validity.

Construct refers to a particular domain such as, for example, Mathematics. For an evaluation method to be construct valid, the tasks it entail, should display or elicit skills and competencies that are construct specific and that proves mastery of the particular construct. Consequential validity on the other hand refers to the ability of a measurement instrument by means of its tasks, to give a trustworthy reflection of intended (as well as unintended) learning outcomes (1994: 100, 101).

With regard to construct validity, a number of criticisms are raised against performance-based measurement. These criticisms include aspects such as: The fact that specific domains are not adequately defined in performance-based measurement; and the fact that performance-based measurement is more task-driven than domain-driven (Gipps, 1994: 100). Gipps suggests that in order to overcome these criticisms, performance-based evaluators should investigate the processes that learners use to perform the instructed tasks (1994:101).

With regard to consequential validity, the proponents of performance-based measurement claim that its tasks do reflect future learning outcomes and that it even promotes the teaching of higher order skills (Gipps, 1994:101; Khattri & Sweet, 1996:16). The criticism against performance-based measurement in this respect is that performance-based evaluators do not normally provide clear conceptual frameworks of the learning that they intend to evaluate. The other danger is also that even higher order skills promotion might become mere rote tasks if the emphasis is too much on the training of tasks.
A further form of validity touched upon by Khattri & Sweet (1996:15, 16) is what they call evidential validity. Evidential validity, according to them refer to the match between measurement tasks and curricular aims. For performance-based measurement to show evidential validity there should be a distinct link between the aims and goals of the curriculum and the tasks that learners are instructed to perform.

Linn & Baker (1996: 89 - 98) discuss the validity of performance-based measurement in relative detail. They distinguish between two categories of validity criteria, namely internal and external criteria.

Under internal criteria they discuss the following aspects:

- "Content quality": This criterion refers to the aim that the content of evaluation methods should be able to improve the instruction process and its targets. Not only must the evaluation activities be worthwhile for both learners and raters, but it must also foster better understanding of subject content and help to eliminate misconceptions. Linn & Baker states that a belief exists among proponents of performance-based evaluation that it does comply with this criterion (1996: 90).

- "Curricular importance": Linn & Baker (1996:91) state that it is possible for performance-based measurement to comply with this criterion provided that it includes tasks which form an important part of the curriculum. The evaluation tasks should therefore focus on the aims and goals of the curriculum.

- "Content coverage": Linn & Baker (1996: 91) argue that performance-based measurement cannot address this criterion in the traditional way (by means of sampling). However, proponents thereof suggest that the fact that performance-based tasks involve multiple processes and the integration of various subject contents, makes it possible for performance-based measurement to cover a wide section of subject content. The only consequence of this approach to content
coverage is that all the performance-based tasks will have to be analysed to
determine its relationship to curriculum aims.

• "Cognitive complexity": Linn & Baker (1996:91, 92) argue that on the surface it
appears as if performance-based measurement complies with this criterion, because
they apparently encourage the testing and utilisation of more complex and higher-
order cognitive processes. However, they reason that the complexity of a task
cannot only be determined by a superficial scrutiny, because certain apparent
higher-order processes can be memorised by learners without them actually
mastering those higher-order cognitive skills. They suggest that it is therefore
necessary that students’ responses and other related aspects should be analysed to
ensure that such higher-order cognitive processes and skills are in fact manifested
by learners.

• "Linguistic appropriateness": Linn & Baker (1996:92) state that certain linguistic
demands which are not essential in achieving the aims of measurement, should be
minimised since it in fact interferes with the measurement process. In terms of this
criterion it appears as if many existing performance-based measurement practices
do not comply since a number of language demands normally accompany them.
Linn & Baker suggest that educators making use of performance-based
measurement should ensure that the linguistic demands set on learners, be dictated
by the goals of instruction that are to be measured.

• "Ancillary skills": Linn & Baker (1996: 92) claim that ancillary skills (that is skills
or outcomes which are not an essential part of the intended outcomes that are
being measured) can pose a serious threat to the validity of any measurement
instrument. Examples of such ancillary skills are speed of response, reading, the
ability to complete the tasks within a fixed period of time, and so forth. Linn &
Baker therefore advise the users of performance-based measurement to minimise
the effect of such ancillary skills in the measurement process by means of task
analysis and the identification of the main intention(s) of the measurement process.
• "Meaningfulness of tasks for learners": Linn & Baker (1996: 93) claim that one of the central reasons for the use of performance-based measurement is the fact that it engages learners in meaningful tasks and activities and this in turn promotes better understanding and motivation among learners. However, the validity of this claim about performance-based measurement still requires some verification.

Under external validity Linn & Baker discuss the following aspects:

• "Consequences for learners and teachers": Linn & Baker (1996: 93) state that an evaluation of the impact of measurement methods on learners is an essential requirement of validation. However, they reason that both intended and unintended outcomes (even negative outcomes) should be examined. In addition to this, they suggest that the influence of measurements on schools, the curriculum and instruction time, should also be evaluated.

• "Fairness": Linn & Baker (1996:94) regard fairness as another essential requirement for the validity of performance-based measurement. However, they make it clear that a change to performance-based measurement will not automatically eliminate deep rooted racial, cultural, linguistic and skills-related differences that exist in societies. They suggest that two main questions should be posed at performance-based measurement, namely:

To what extent does performance-based measurement cater for various cultural backgrounds of learners?

What differences are there in terms of opportunities to learn and master the most essential skills that are intended in the instructional process?

In conclusion it would appear that performance-based measurement has the potential to comply with all of the validation requirements discussed above. However, special attention should be paid to areas where there are some shortcomings. Such special attention should include the training of raters and careful analysis of tasks and its relationship to curricular aims.
4.4.5 Reliability of performance-based measurement

Gipps (1994:103, 104) states that one of the main aspects that is focused on with regard to the reliability of performance-based measurement, is the aspect of inter-judge/rater agreement in the scoring process. Burry-Stock et al (1996: 254) state that inter-rater agreement has to do with the extent of agreement between judges on the rating of a particular behaviour.

According to Gipps, inter-rater agreement has become an important factor due to the subjective nature of marking in performance measurement and the fact that the process relies heavily on professional judgements. She contends that inter-rater agreement can be very high provided that raters are carefully trained and supplied with scoring rubrics. Khattri & Sweet (1996:17) also emphasised the need for sufficient training of raters and the standardisation of rating criteria. Gordon, Engelhard, Gabrielson & Bernknopf (1996: 85) however issue a cautionary note with regard to rater training for the purpose of ensuring inter-rater agreement. They state that despite rigorous training programmes, and possibly a certain level of consensus, raters have certain “ingrained response tendencies” that will continue to exist. These ingrained personal preferences manifest itself in either more lenient or more stringent approaches by different raters.

In terms of consistency of performance across tasks or score reliability, performance-based measurement appears to be performing poorly (Gipps, 1994: 104, 105). It appears that performance on different tasks is not highly related, but rather task specific. Gipps (1994:104) and Khattri & Sweet (1996:17) advise that in order to improve score reliability, that the number of tasks should be increased in every evaluation exercise.

Linn & Baker (1996:95) also point out the aspects of inter-rater agreement and the limited level of across-task generalizability. They also advise the increase of tasks in order to overcome this shortcoming (low task generalizability).
In conclusion it would appear as if in terms of reliability, performance-based measurement has the potential to comply with the set requirements, but that there is still much room for improvement, especially with regard to generalizability. The onus is thus on proponents of performance-based measurement to ensure that enough tasks are designed to ensure that the shortcoming of low inter-task generalizability is overcome. With these few comments, the aspect of accountability of performance-based measurement will be addressed next.

4.4.6 Accountability of performance-based measurement

The issue of accountability appears to be a major one with regard to evaluation/measurement instruments. Some scholars feel that performance-based measurement complies very well with this crucial measurement requirement because of its direct and skills-based form of measurement, while others feel that because of the absence of national standardisation, that it lacks very much in terms of accountability.

Hustler (1995: 116, 117) states that educational institutions are accountable to potential employers and the public concerning the education and training of learners or potential employees. For this purpose he advises a strive towards quality service and products by all involved in the education process. He further states that evaluation and the supply of feedback to learners about their academic progress and performance were identified as some of a few indicators of quality in higher education about which there is uniform agreement. Another aspect which especially reflects positively with regard to performance-based measurement, is the fact that active student involvement in the learning/evaluation processes is another of the few quality indicators about which unanimity exists in England. Hustler (1995: 117) also mentions the growing demand that learners should take greater responsibility with regard to their own learning and progress. This is also a factor that features strongly in performance-based measurement where especially the aspect of student self-responsibility and empowerment is strongly promoted (for example self-evaluation, portfolios, profiles, and so forth).
Resnick & Resnick (1996: 34) distinguish between two models of accountability in education. The first model is one based on a traditional one in the US which displays the following characteristics: The different states supply a set number of resources and ensure the establishment of minimum standards for passing certain courses, while districts are concerned with the employment of teachers and administrators and schools are concerned with the curriculum. The assumption is made that as long as there is no major upsets in this status quo, there is a built-in type of accountability present. This model is also characterised by passivity of most of the essential role-players (learners, teachers and parents). Measurement instruments in which very little activity or performance is required will suit this accountability model.

The second accountability model distinguished by Resnick & Resnick (1996: 34) is the total opposite of the first one. In this model there is a joint active involvement in the total learning process by learners, parents and teachers. There is direct communication about learning aims and objectives and ways to manifest the achievement of those aims and objectives. Resnick & Resnick suggest that performance-based measurement will perfectly blend in with this model, because this model requires direct measures of actual skills and performance of learners. Resnick & Resnick argue that such direct measurement should be a central factor in the establishment of an accountable evaluation system.

Hall (1989: 25) states that the introduction of records of achievement as a form of evaluation/measurement, was a result of a growing demand in society for greater accountability on the side of educational institutions. This accountability is demanded from schools and teachers by parents, learners and potential employers. Hall stresses the fact of an increasing demand for accountability which is probably partly a result of the former poor state of accountability in educational institutions.

Hall (1989: 25) further states that the attempts towards greater accountability by means of records of achievement was initially placed in the lap of individual teachers and schools. However, criticism soon developed stating that this form of reporting did not provide a consistent form of accounting to employers. The issue of credibility of
records of achievements was also raised. The essence of this criticism was that national principles and standards (standardisation) should be incorporated into records of achievement. However, it would appear as if records of achievement do comply with certain accountability requirements (such as accountability to parents, to learners, to the wider community, and so forth) mentioned by Hall (1989: 26) because it gives comprehensive information about students' performances, achievements and skills.

Smith & Levin (1996: 117) argue that before the accountability of performance-based measurement can be determined, the central issue of an accountability system should first be sorted out. However, they mention a number of obstacles in the way of establishing an accountability strategy based on performance-based measurement/ measurement.

The first of these obstacles is the issue of legitimacy. They state that it would be very difficult to introduce a performance-based accountability strategy that has the mandate to punish both the educational institution and the learners if they do not comply with its basic requirements. This difficulty arises from the fact that there is no unanimity with regards to who should be accountable to who. Some stakeholders feel that an accountability system should be founded on simultaneous co-responsibility between institution and learners, while others feel that accountability should start at the institution first. The latter group argues that a student's performance in a particular test or examination can be influenced by many factors, for example good/poor teaching, adequate resources, teacher support, and so forth. The dilemma is described very accurately by Smith & Levin:

“But if learners cannot be held accountable until their school is held accountable, and no school can be held accountable unless its learners are accountable, and neither can be held accountable until schools have the necessary resources to succeed, legitimising high stakes assessments becomes a very complex endeavour.” (Smith & Levin, 1996: 117)

The second major obstacle mentioned by Smith & Levin (1996: 118) is the drive among educational institutions towards the achievement of high pass-rates. Smith &
Levin argue that this competition leads to short term “beat the test” approaches which do not serve any long term aims regarding the instruction process or the curriculum. It can also lead to a miss-directed type of accountability, because the focus does not really fall on giving an account to stakeholders about students’ achievements and skills, but on own prestige. They reason that performance-based measurement can at best lessen the severity of this competition factor, but not totally erase it.

Smith & Levin (1996: 118) further argues that the basis of any accountability in an educational institution is a solid and well functioning administration system. They reckon that without such a sound administrative system, even the best functioning evaluation system will be of no use. They feel that the essential requirements of an accountable evaluation system is that it should be able to contribute towards improved teaching and learning. The evaluation methods used, should give learners clear direction of their own progress and mastery of required skills. When an analysis of various types of performance-based measurement instruments is undertaken, it appears as if this is exactly what is aimed at in performance-based measurement/measurement. Khattri, Kane & Reeve (1995: 80) for example claim that there is growing evidence that performance-based measurement does contribute strongly to the improvement of both teaching and learning. Based on such information it becomes clear therefore that with regard to this form of accountability, performance-based measurement will pass the test.

In conclusion it would appear that while the debate of the extent of accountability of performance-based evaluation/measurement still rages on, there are certain definite areas in which performance-based evaluation/measurement does comply with the requirements of accountability (for example giving a clear account to learners, parents and possible employers of the academic progress and achievement(s) of learners and the skills that they have mastered in the learning process). As soon as the aspect of standardisation can be sorted out, even less criticism against the accountability of performance-based measurement can be raised. With these brief comments the following aspect that will be examined is the general or common advantages of performance-based measurement/measurement.
4.4.7 General advantages of performance-based measurement

Gipps (1994: 102 - 105) mentions a number of aspects that can be regarded as advantages of performance-based measurement. Such aspects include the following:

- The fact that performance-based measurement allows for direct observation of skills and competencies to be measured.
- Performance-based measurement is generally regarded as a more professionally acceptable form of educational measurement.
- The validity and reliability levels of performance-based measurement can be increased if the correct steps are followed.

Rivera et al (1995: 34) also mention a number of aspects regarding performance-based measurement that can be regarded as advantages thereof. These aspects are:

- It promotes not only the development of critical thinking skills in learners, but it also facilitates student self-reflection on their own work, knowledge and skills.
- It makes possible the practising of real-world skills by means of the authentic activities it requires.

Khattri et al (1995: 80 - 82) mention the following advantages of performance-based measurement:

- Firstly they claim that there is growing evidence that performance-based measurement does contribute strongly to the improvement of both teaching and learning (as mentioned in section 4.4.6).
- Secondly they state that performance-based measurement positively contributes towards student motivation.
- It also helps to improve students’ writing skills.
To summarise, it would appear as if there are certain definite advantages involved in the use of performance-based measurement instruments. Especially the aspects regarding meaningful tasks and real world skills that are tested, are very important improvements on the conventional measurement instruments as discussed in chapter 3. However, despite these profound advantages of performance-based measurement, there are certain undeniable disadvantages attached to it which will be discussed in the next section.

4.4.8 General disadvantages of performance-based measurement

Gipps (1994: 102 - 105) mentions the following aspects that can be regarded as disadvantages of performance-based measurement:

- Because of the time-consuming nature of performance-based measurement, breadth of coverage often has to be sacrificed for depth of coverage.
- Performance-based measurement is extremely subjective and judgmental in nature with regard to marking or scoring.
- Task performance is not very generalizable because performance is very task-specific.

Khattri & Sweet (1996: 17) also mention the problem of low inter-task generalizability with regard to performance-based measurement. They further mention the aspect of high costs involved in performance measurement (1996: 18), but they argue that the cost aspect might be misleading, because it is not very easy to determine the exact costs of the total evaluation process. They also suggest that the professional and academic benefits of performance-based measurement should be weighed up against the cost aspect.

Other scholars that emphasise the aspect of high costs include Strong & Sexton (1996: 102) and Kamen (1996: 861). However, although so many scholars are convinced about high cost implications of performance-based measurement, it would appear as if the actual cost implications are not really that phenomenal. Monk (1995: 364) for
example undertook an analysis of the cost implications of performance-based measurement on a state-wide basis in a number of states in the US. He found that although the costs ran into millions of dollars, it represented less than 1 percent of the total spending on elementary and secondary schools in those states. If this is indeed the case in other contexts and in view of the benefits of performance-based evaluation, it would appear as if performance-based measurement might not actually be such an expensive venture as is the perception of many scholars.

With this brief discussion of the possible common disadvantages of performance-based measurement, a more detailed examination of specific performance-based measurement instruments will be undertaken next.

4.5 A BRIEF SCRUTINY OF PERFORMANCE-BASED MEASUREMENT INSTRUMENTS

In the sections that follow, a brief scrutiny of some of the main and most common types of performance-based or alternative measurement instruments will be provided (see section 4.4.2 for the list of instruments). No claim is made that the topic has been exhausted, however quite a wide range of performance-based measurement instruments is covered in the remaining part of this chapter. There will be extensive reference to Priestley (1982) because he provides the most comprehensive account of the various performance-based measurement instruments that the researcher could get hold of.

The first category of performance-based measurement instruments that will be examined is actual performance measurement.

4.5.1 Actual performance measurement

Priestley (1982: 25) states that actual performance measurement involves some sort of real performance by the candidates involved. Actual performance measurement normally takes place by means of observation of individual candidates by the
evaluator(s). Priestley further states that the inevitable subjective nature of observational judgement can be countered by the use of well constructed and objective measurement instruments.

The main advantage of actual performance measurement is the fact that it probably provides the most realistic evaluation of work-related skills and competencies. However, the main disadvantages are the fact that performance evaluation is very time-consuming and a very expensive process in terms of design, administering, equipment required and scoring (Priestley, 1982: 25).

With these introductory remarks a more specific examination of various actual performance measurement instruments will be undertaken.

4.5.1.1 Work-sample tests

4.5.1.1.1 Description

Work-sample tests in essence entail the assignment of a work-related task to an examinee and a judging of the result by means of checklists or rating scales. It can be used for both process- and product-based measurement. In the case of process-based measurement, judgement is based on observation, while in the case of product-based measurement, judgement takes place after completion of the task (Priestley, 1982:26).

McBeath & Lassen (1992: 294) reckon that work sample tests can also be used as a type of role-playing test (see section 4.5.2.4.1) in which actual tasks or problems relevant to a certain type of job are simulated. They use the example of the “in-basket” test which is normally used to simulate administrative tasks and decisions. Learners are normally presented with memos, letters, reports and so forth in which problems occur. Learners then have to demonstrate how they will tackle those problems and what choices they will make.

Examples of areas wherein work-sample tests are frequently used are the following:
- **Product evaluation**
  
  Food preparation (for example a loaf of bread, a completed meal)
  
  Office skills (for example a typed letter or transcript)
  
  Industry
  
  Plastic arts (sculpture, film-making)
  
  Construction
  
  Manufacturing

- **Process evaluation**
  
  Vehicle operation (a road test required for a driver's license)
  
  The performing arts (auditions for performers)
  
  Physical fitness (actual exercises)
  
  Laboratory research (actual experiments)
  
  Sales
  
  Personnel
  
  Office skills (Priestley, 1982: 26).

Although, according to Priestley (1982: 26) many of these measurements take place in an informal manner, it can be used for formal measurement purposes provided that the "mode of assessment, mode of response and method of scoring" are well defined and properly implemented.

4.5.1.1.2 Advantages

Priestley (1982: 71, 72) identifies the following advantages of work-sample tests:

- Work-sample tests exhibit high face validity because it measures actual performance in realistic situations.

- Because of its realistic nature, it is able to measure skills that cannot be easily measured by other methods.

- Work-sample tests make direct observation of a particular performance possible.
• It also makes it possible for the observer (if necessary) to provide constructive feedback to the examinee during or after completion of the task.

4.5.1.1.3 Disadvantages

Priestley (1982: 72) identifies the following disadvantages of work-sample tests:

• They require one-to-one administration and as a result cannot easily be administered to great numbers of examinees at the same time.
• They require a lot of equipment (for a single task) making it very expensive.
• Work-sample tests are very time-consuming.
• They are only feasible in settings where the required equipment is normally available or where specific apprenticeship programmes are in operation.
• Work-sample tests are not always standardised and many of the examinees might not be familiar with the equipment used for testing.
• Observation in work-sample tests tend to be highly subjective which decreases its reliability.
• Work-sample tests can only be best used in areas or tasks that can be easily reduced to their simplest components and therefore excludes more complex types of tasks (for example management, supervision, psychiatric counselling, sales, and so forth).
• Because of cost and time limitations, work-sample tests are in effect only samples of a certain task or tasks and cannot be used to measure complete domains of required skills.

4.5.1.1.4 Suitability for TSA conditions

In the traditional distance education setting, where learners had to cope mainly on their own, and lecturers had to mark large numbers of exam scripts, the use of work-sample tests at TSA would be very impractical, if not impossible to administer because of the disadvantages of work-sample tests (see section 4.5.1.1.3).
However, with the new decentralised approach and tutor-system adopted by TSA since 1995, these hurdles can be overcome to a great extent because of the following reasons:

- The intention is that tutors will be assigned to smaller numbers of learners (1 - 50) making that one-to-one interaction between student and tutor possible.
- The smaller number of learners per tutor will also make it possible that more time can be spent on individual learners.

The only serious problem remaining, is the one of getting the required equipment, which will probably (because of the costs involved) be a very difficult if not impossible aim to achieve for an educational institution such as TSA. A possible solution to this problem might be to liaise with industries who own such equipment with the aim of them availing their equipment for educational evaluation purposes. However, this option might not succeed everywhere and it will need a lot of prior research.

4.5.1.1.5 Suitability for BM

In view of one of the disadvantages of work-sample tests as mentioned in section 4.5.1.1.3, namely the fact that it can only be used effectively for subjects or tasks that can be easily reduced to their basic components, it would appear as if work-sample tests might not be very suitable for Business Management. This contention is based on the fact that BM is essentially a management-based subject and its actual skills and tasks cannot be simplified readily.

4.5.1.1.6 Financial implications

It has already been pointed out in section 4.5.1.1.3 that work-sample tests will in effect be a very expensive measurement instrument because of the great quantity of essential equipment that will be required as well as its time-consuming nature. Work-sample tests might therefore be very difficult to implement at TSA which, similar to all
other educational institutions in SA, is forced to search for more cost-effective evaluation methods.

4.5.1.2 Identification tests

4.5.1.2.1 Description

In its simplest form, an identification test requires from examinees to identify something which may range from a tool, object, material, problem or function to a quality in a particular job or performance (Priestley, 1982:26). Identification tests are regarded as actual performance tests because they require the identification of actual objects or problems in a realistic setting (Priestley, 1982: 75). However, the identification test has frequently been used for more than mere identification purposes. Priestley (1982:26, 27) states that identification tests can also refer to actual performance, simulation and even paper and pencil tests.

In an actual performance for example, the examinee might be required both to perceive or identify a particular problem or malfunction and to respond or try to fix it. The examinee is in the first place required to use one or more forms of perceptual stimuli which can be auditory (listening), gustatory (tasting), tactile (feeling), olfactory (smelling), or visual. The examinee can respond in a variety of ways (depending on the nature of the test) which can be oral, written, motion or particular actions. A mechanic might for example be required to locate a problem in a certain engine and fix it. Identification tests are frequently used in combination with work-sample tests (Priestley, 1982:27).

Examples of where identification tests are regularly used are:

- Screening tests for entry-level examinees, for example:
  - carpentry
  - nursing
- Occupations where identification forms an essential skill, for example:
- anaesthetists (gases)
- chemists (gases)
- chefs (spices)
- air traffic controllers (types of planes)

- For diagnostic or "troubleshooting" skills, for example
  - to identify a problem or malfunction
  - auto mechanics
  - television repairs
  - home economists
  - veterinarians (Priestley, 1982: 76).

4.5.1.2.2 Advantages

Priestley (1982:77) identifies the following advantages of identification tests:

- The main advantage is the fact that it can adequately assess candidates' diagnostic skills and their ability to identify problem areas in a job setting.
- They can be used for the measurement of perceptual competencies.
- It is very realistic and direct because the examinee has to identify real objects or substances which occur in the actual work set-up.
- They are very useful for the establishment of the entry-level competencies of candidates.
- It can be used for small groups of examinees at the same time and is therefore not limited to single individuals.

4.5.1.2.3 Disadvantages

- Such tests can be very expensive to conduct because it requires the use of expensive or difficult to locate material, tools or objects.
- It only measures basic knowledge or simple skills.
- It does not directly measure actual performances (in the case of mere identification).
• Identification of a particular malfunction does not automatically imply the skill to fix such a malfunction.

• Because of these disadvantages identification tests are mostly used for entry-level purposes or in combination with other tests, for example work-sample tests (Priestley, 1982:77).

4.5.1.2.4 Suitability for TSA conditions

More or less the same comments under 4.5.1.1.4 will also be applicable here. The only difference is that identification tests can be used for more than one individual at a time as indicated in section 4.5.1.2.2, but still not for very large numbers of learners at the same time.

4.5.1.2.5 Suitability for BM

It would appear as if identification tests can be used in BM for measurement purposes, but to a very limited extent. As mentioned in section 4.5.1.2.3, identification tests are normally only used to measure basic knowledge and skills in a particular job or subject, and would therefore only be used for such basic knowledge and skills contents in BM. It can also be used to establish a student's entry-level when enrolling for the subject BM, or at the time of applying for a job at a particular industry.

4.5.1.2.6 Financial implications

It seems as if identification tests can also be regarded as a very expensive measurement instrument because of the costs of the material, tools or equipment required (Priestly, 1982:77).
4.5.1.3 Supervisor-ratings

4.5.1.3.1 Description

Supervisor-ratings can be regarded as merely another way of assessing various aspects of an employee's work performance (Priestley, 1982:82). Supervisor ratings make possible the unobtrusive observation of employees while they are engaged in their normal work activities. Both product and process can be measured in this way. Supervisor-ratings do however require the setting of predetermined criteria and the recording of employees' performances on the basis of that criteria. The supervisor ratings are based on certain work-related attributes (for example performance, behaviour, and so forth) which should be identified by means of a job-analysis, and should not include unrelated opinions or attitudes. Instruments such as checklists, rating scales or anecdotal records are normally used for supervisor ratings (Priestley, 1982: 82).

Supervisor ratings are normally used for the continuous assessment of employees' progress in their jobs for future promotion or salary increase purposes. It can also be used to test people in apprenticeship or internship programmes. Supervisor ratings can also be used both for the evaluation of the products produced by employees, as well as their actual performance on the job (Priestley, 1982: 83).

4.5.1.3.2 Advantages

Priestley(1982: 83) lists the following advantages of supervisor ratings:

- Firstly, it embodies most of the advantages inherent in observational assessment.
- Secondly, supervisor ratings make possible the direct measurement of skills and behaviours.
- It is used only to rate skills and behaviours that are work-related.
- The personal comments or recommendations by supervisors can improve the appraisal process of candidates.
• Because of the fact that supervisors and the evaluation set-up are familiar to the employees, they will probably feel less anxious during the process of evaluation.

• Supervisors can provide employees with constructive feedback in the process which might lead to improved performance.

4.5.1.3.3 Disadvantages

Priestley (1982:83) identifies the following disadvantages of supervisor ratings:

• Bias, for or against candidates, appear to be inevitable because different candidates are normally rated by different supervisors.

• It is not a very easy process to establish the criteria for evaluation because supervisors must ensure that it is only relevant, job-related skills and behaviours that are to be evaluated.

• There might be a lack of reliability and consistency in the scoring and interpretation processes because of the inevitable subjective nature of the rating process.

• Because of own job responsibilities, supervisors are not always available for such ratings which makes the establishment of a standardised rating process very difficult to achieve.

• Unsystematic supervisor ratings (which is frequently a normal consequence of the above disadvantage) are not regarded as valid and reliable forms of evaluation.

4.5.1.3.4 Suitability for TSA conditions

In as far as TSA's distance education practice in general is concerned, supervisor ratings will not be possible, because the learners are not necessarily employed in an institution where on-the-job evaluation by actual supervisors can take place. There are however three possibilities in as far as the decentralised approach at TSA is concerned:

• Firstly, the tutors can act as supervisors at the normal contact sessions that will be offered at the study centres. However, this approach will be too theoretical and will not allow room for the display of actual job-related skills and performance.
• Secondly, if practically possible, arrangements could be made with main industries in the main centres of different provinces where TSA branches exist, to let learners undergo internship or apprenticeship programmes at such industries and letting them be rated by actual supervisors.

• Finally, if industries can be co-opted into this process, but cannot make their supervisors available for such additional ratings, tutors can be used to rate learners that are placed in internship or apprenticeship programmes at particular industries.

In conclusion, it would appear as if to effect supervisor ratings for learners, TSA should consider embarking on internship/apprenticeship programmes for learners.

4.5.1.3.5 Suitability for BM

Most of the comments in section 4.5.1.3.4 will also apply in this instance.

4.5.1.3.6 Financial implications

It would seem as if the financial implications of such internship/apprenticeship programmes might be enormous. However, if industries are willing to assist TSA as an educational institution in the availing of their premises, equipment, and staff for such programmes, while TSA can use tutors for the actual supervisor ratings, it might be possible to implement supervisor ratings as a form of evaluation in BM but preferably only on final year level.

4.5.1.4 Peer ratings

4.5.1.4.1 Description

Peer ratings involve the evaluation of candidates’ work or performance by their colleagues or peers. According to Priestley (1982:85) peer ratings are known as “peer reviews” in professional circles. In the case of the medical profession for example, such peer reviews are very highly acclaimed, so much so that it even serves as the main
basis of evaluation (Priestley, 1982:85). The criteria used in peer ratings would be similar to those used in supervisor ratings.

Peer ratings are normally used in the following instances:

- In training programmes to:
  - improve/enhance the education process;
  - facilitate self-evaluation by trainees;
  - promote self-awareness and personal growth.
- It is normally used for the evaluation of (final) products because the evaluation of processes are many a time not observable or peers might not be qualified enough to evaluate procedural techniques.

4.5.1.4.2 Advantages

Peer ratings will normally have the same advantages as supervisor ratings. Priestley (1982:85) suggest a few additional advantages, namely:

- Evaluation by peers might be less threatening to trainees than supervisor ratings.
- The participants (evaluators) also benefit because they learn more about themselves and the total evaluation process.
- Peer ratings will probably also be cheaper to implement than supervisor ratings.

4.5.1.4.3 Disadvantages

Peer ratings also involve more or less the same disadvantages as supervisor ratings (see section 4.5.1.3.3) with a few additional ones identified by Priestley (1982: 85, 86).

- Peers do not necessarily know better than the persons they have to evaluate what job-related skills and behaviours are essential and whether the particular person has mastered them.
• Peer ratings suffer from a very low standard of reliability and validity because the raters are not trained evaluators and they will be evaluated by the same standards.
• A factor that might also negatively effect the validity of peer ratings, is the fact that peers might experience a feeling of competition with those that they have to evaluate.

A few additional disadvantages that can be foreseen by the researcher, are:

• Peers might feel obliged to their close friends and might not give an accurate and objective rating.
• Peer ratings might be very time-consuming in cases where groups of learners have to evaluate all the members in a group.
• Peer ratings in instances where no internship programmes are in operation will merely be based on the theoretical side of the subject (the information in the study material).

4.5.1.4.4 Suitability for TSA conditions

In cases where large numbers of TSA learners live in the same areas or close to the same study centres, peer ratings can be implemented. It can also be a very useful form of evaluation for distance education where learners, despite the availability of study centres, remains dependent on their own mastery of study material. Peer ratings can in fact serve as a means of encouragement and motivation of learners among themselves, provided that it is conducted in a positive and constructive spirit. Peer ratings will in fact be more easily implemented and cheaper than supervisor ratings. However, peer ratings should only serve as an additional form of evaluation because of some of its disadvantages.

4.5.1.4.5 Suitability for BM

See 4.5.1.4.4.
4.5.1.4.6 Financial implications

As mentioned in section 4.5.1.4.2, peer ratings will be a much more cost-effective measurement instrument, because learners would not normally demand any financial benefits for evaluating one another. Peer ratings is in fact one of the most cost-effective forms of evaluation (even more cost-effective than written examinations).

4.5.1.5 Self-evaluation

4.5.1.5.1 Description

Self-evaluation is a form of evaluation that requires learners to evaluate their own performance in whatever job or subject. Self-evaluation can take the form of introspective observation, assessment of own products or own testing which will be self-administered without any outside control. Self-evaluation is used more as a tool for instruction than for testing (Priestley, 1982:28).

Boud (1991:5) regards the following two aspects as the central pillars of student self-evaluation:

- The identification of standards or criteria which learners will apply to their own work.
- Learners judging to what extent they have achieved those criteria or standards by means of self-reflection.

Boud also argues that self-evaluation is not merely another form of testing, but a process of enabling learners to become involved in and take responsibility for their own learning. Such self-responsibility or self-accountability will inevitably lead to the empowerment of learners which will enable them not to be solely dependent on educators for their own learning and development, but to function independently.

Techniques used for self-evaluation can be formal or informal. Formal techniques include self-tests, structured interviews, checklists and rating scales, while informal techniques can take the form of discussions with a supervisor or teacher. Self-
evaluation can be used for both product-based and process-based evaluation. In order to effect self-evaluation the candidate should be properly instructed with regard to the "how" of self-evaluation and the use of suitable criteria (Priestley, 1982:86).

The main aim of self-evaluation is to help learners learn more effectively and can therefore probably be used in virtually all subject areas. The three main self-evaluation instruments and their uses are stated below:

- Self-evaluation tests can help to give learners or examinees a structured account of the extent of their learning in a particular area.
- Inventory checklists and job information forms can be used for the purpose of screening job applicants, identifying the critical features of any job, or to evaluate training programmes.
- Performance checklists or rating scales can help learners to evaluate own performance in a structured way (Priestley, 1982: 86, 87).

4.5.1.5.2 Advantages

Priestley (1982: 87) identifies the following advantages of self-evaluation:

- Learners are enabled to recognise own strengths and weaknesses.
- Learners are trained how to use evaluation criteria and how to undertake objective judgements.
- Learners are shown the value of well defined guidelines, quality control measures and specific procedures.
- Self-evaluation, if learners are well acquainted to it, is probably the least threatening form of evaluation compared to teacher, supervisor of peer evaluation.
- Another advantage that the researcher can foresee (as mentioned earlier in this section), is the fact that self-evaluation can serve as a powerful tool for the empowerment of learners.
4.5.1.5.3 Disadvantages

The following disadvantages of self-evaluation are indicated by Priestley (1982: 87, 88):

- Learners normally suffer from a lack of experience in conducting evaluation which makes the implementation of self-evaluation very difficult in its initial stages.
- Self-evaluation frequently appears to be inflated and lacking in objectivity as a result of self-interest and emotional involvement.
- Probably the main disadvantage of self-evaluation is the fact that it cannot serve as an official form of evaluation with legal consequences as a result of its lack in objectivity and reliability.

4.5.1.5.4 Suitability for TSA conditions

Self-evaluation is ideally suitable for distance education because of the fact that learners in distance education are mainly reliant on themselves for understanding, interpretation and mastery of study material. Self-evaluation is therefore a very effective method to help learners measure the extent to which they have mastered or grasped certain sections of the work. The main requirement is that learners should be well trained in the use of self-evaluation so that they can use this technique with optimal effectiveness. However, unless learners are in an actual work setting, the self-evaluation will be merely based on theoretical subject knowledge from the study material and textbooks.

4.5.1.5.5 Suitability for BM

Self-evaluation, as mentioned in section 4.5.1.5.1, can be used in virtually all subjects and will therefore also be suitable for BM. However it will also merely be able to serve as an additional form of evaluation because of its subjective nature and the fact that it is a non-standardised form of evaluation/measurement.
4.5.1.5.6 Financial implications

Self-evaluation is even a more cost-effective form of evaluation/measurement than all the others discussed so far and almost no costs need to be incurred. One great advantage of self-evaluation above peer ratings, is the fact that it is also very effective in terms of time because it is only one candidate that has to be evaluated. This aspect can help to achieve an even more comprehensive and higher quality of evaluation.

In conclusion to this section regarding actual performance measurement, the following observations can be made: It appears as if actual performance measurement is an ideal type of measurement because of its direct nature. However, actual performance measurement will be the most suitable in instances where learners are in actual work settings. For actual performance measurement to be implemented at TSA will inevitably require the introduction of apprenticeship/internship programmes (see section 4.5.5.3) There seem however to be very serious obstacles in the way of implementing actual performance measurement. These include aspects such as high costs, subjectivity in the scoring process and the lack of standardisation.

With these brief comments the next category of performance-based measurement instruments will be examined next.

4.5.2 Simulations

Simulations are normally used in cases where actual performance tests are impossible or very impractical because of reasons such as cost, danger, the serious consequences in the case of mistakes, and so forth (Priestley, 1982: 91). Simulations are designed to be as close to real life settings as possible and therefore the settings and tasks to be performed in a simulation, should be as realistic as possible. The greatest value of simulation is the fact that it makes it possible to predict fairly accurately how well candidates will perform in real situations. Simulations are not merely used for the purpose of evaluation, but rather more frequently as training or teaching tools. One of the advantages of simulations above real performance is the fact that examiners are
able to control most of the variables in the testing situation (which is not possible in actual performance testing). This factor makes possible the standardisation of simulations as a form of measurement on a national basis (Priestley, 1982: 91).

McBeath & Lassen (1992: 294) state that most performance-based measurements are conducted in simulated settings. Such simulation settings can include aspects such as material, equipment, instruments and so forth, and other people who help to create a realistic setting.

The main forms of simulations that will be examined are:

- Simulated performance tests
- Simulated identification tests
- Written simulations
- Management exercises
- Simulator machines

4.5.2.1 Simulated performance tests

4.5.2.1.1 Description

A simulated performance test involves actual physical performance, but in a simulated setting. In other words, the performance itself will be exactly the same as in the case of an actual “on-the-job” situation, but the actual setting will not be an actual job setting, merely a simulated one (Priestley, 1982:29, 92). Simulated performance tests are normally used instead of actual performance tests in cases where the consequences of mistakes might be too serious (for example surgical operations, tunnel blasting, assembling of explosives, and so forth).

Simulated performance tests are used for purposes such as:

- Examinees’ competence in technical procedures
- To evaluate management skills
Simulations can in fact be used in a great variety of areas and subject disciplines. A well-known example of simulated performance tests and which deserve special attention, is the so-called “in-basket tests” (Priestley, 1982: 94). This test normally simulates the administrative tasks of a management executive and the name “in-basket” refers to the different baskets (for example incoming information, outgoing information, and so forth) normally found on the desk of managers or executives. In the case of an in-basket test, the examinee will in the first place receive background information about the job that has to be performed. Following this, examinees will be instructed to perform that particular job and react on every new document that lands into one of their baskets. Two main tasks are required from examinees, namely to demonstrate the course of action they will follow after receiving new documents (for example file it, write a letter, inform the executive, telephone someone to supply certain information, and so forth) and to state the reasons for the actions that they have taken. All these courses of action and reasons will be submitted at the end of the test for evaluation purposes, but during the course of the in-basket test, examinees will also receive points for taking certain actions or the lack of any actions (Priestley, 1982: 94).

This type of simulated performance tests can be very useful in business settings.

4.5.2.1.2 Advantages

Priestley (1982: 30, 97) identifies the following advantages of simulated performance tests in general:

- They allow for realistic and relatively direct evaluation of on the job skills and behaviours.
- It makes predictions of actual performance in actual settings possible.
- They serve as effective methods for training learners and preparing them for actual realistic situations.
• Because of the fact that the examiner can control most of the variables in the simulation, it is possible to standardise tests across all examinees and testing situations.
• This standardisation improves the reliability of such simulated performance tests.

4.5.2.1.3 Disadvantages

• According to Priestley (1982: 97) the main disadvantages of simulated performance tests are related to costs and time requirements. The in-basket tests for example require a lot of instruments and equipment and require up to two days to administer and score properly.
• Another disadvantage is that the scoring procedures are very complex, involved and a lengthy process.
• A third disadvantage is that it is less realistic than actual performance testing (Priestley, 1982: 30).

4.5.2.1.4 Suitability for TSA conditions

Simulated performance tests appear to be quite appropriate for distance education, especially because TSA as an educational institution might not be able to secure agreements with companies in all the areas where they have study centres for the establishment of internship or apprenticeship programmes (as advised in section 4.5.5.3). Simulated performance tests are therefore very useful in cases where actual on the job settings are not possible to obtain. It will also be more suitable in the case of small numbers of learners because of the time-consuming nature of for example in-basket tests. However, the new decentralised tutor system of TSA can make this form of evaluation/measurement possible if the intended 1 - 50 tutor/student ratio can be established. The only major problem that might arise is with regard to the cost factor because to get hold of the required material, equipment and other relevant aspects, will probably be a very costly process. But once again, the advice will be to liaise with industry and to get their support in terms of sponsoring the evaluation process and making their facilities available to TSA.
4.5.2.1.5 Suitability for BM

Especially the in-basket tests appear to be a very suitable form of evaluation for BM because the subject does involve one of the types of careers that were illustrated in section 4.5.2.1.1 (for example warehouse managers, purchase managers, dispatch managers, and so forth). It will also help to lessen the difficulty experienced in BM to clearly define the skills that learners should actually master.

4.5.2.1.6 Financial implications

As already mentioned in section 4.5.2.1.3, the financial implications of simulated performance tests can be enormous. Acquiring all the necessary material, equipment and instruments for all TSA centres will be a very expensive venture. The only possible solution to avoid these high costs would be to obtain the support and assistance of industry.

4.5.2.2 Simulated identification tests

4.5.2.2.1 Description

In a simulated identification test examinees are required to identify certain parts, certain problems or to manipulate a model (for example a mannequin), but nothing more. The test situation is a controlled one and is therefore not in an actual work situation. The examinee is not required to identify real problems or materials and does not perform an actual function in a real situation. Simulated identification tests are mostly used to measure cognitive understanding of how to perform certain tasks, rather than the actual performance of tasks (Priestley, 1982: 30, 100).

Simulated identification tests are normally used in the following careers and situations:

- In cases where mistakes in the real situation will have too serious consequences.
• In cases where it would be impractical to reproduce malfunctions for every student (for example a machine working at the wrong speed).
• For medical laboratory technicians.
• Engineering
• Architecture
• Construction, and so forth (Priestley, 1982: 100, 101).

4.5.2.2.2 Advantages

Priestley (1982: 101, 102) points out the following advantages of simulated identification tests:

• Although less direct than actual performance tests, the simulated identification test is more direct than paper-and-pencil identification tests.
• It is less expensive than actual performance tests (but it remains an expensive process).
• The examiner is in a position to control most variables in the test situation making possible the standardisation of the full procedure.
• The serious consequences of mistakes are avoided in the case of simulated identification tests.

4.5.2.2.3 Disadvantages

The following disadvantages are identified by Priestley (1982:101, 102):

• Simulated identification tests are limited to the evaluation of cognitive ability and does not measure actual skills or performance.
• Although the equipment and materials required might be less expensive than in the real situation, it still remains a very expensive exercise.
• The accuracy, completeness and functioning ability of simulated materials might not be on the same level as that of the real instruments.
4.5.2.2.4 Suitability for TSA conditions

Most of the comments in section 4.5.2.1.4 will also apply to simulated identification tests.

4.5.2.2.5 Suitability for BM

When BM is compared with the other subjects and areas mentioned in section 4.5.2.2.1 for which simulated identification tests are normally used, it appears as if the nature of BM as subject differs to a great extent. Although there might be cases of problems or malfunctions, it will be different than in the case of the subjects mentioned. Also the fact that it focuses primarily on the cognitive domain will make it less desirable, because this research is precisely attempting to find alternatives to cognitive-based forms of measurement. It would therefore appear that although simulated identification tests can be used in BM, it is not one of the most recommended or useful forms of measurement for BM. When the cost factor is also taken into consideration, it is questionable whether the use of simulated identification tests in BM will be worth the costs.

4.5.2.2.6 Financial implications

The cost implications of simulated identification tests were mentioned in section 4.5.2.2.3 and based on that, it would appear as if the high cost factor and the small extent of usefulness with regard to BM will not make this form of measurement a very effective one for BM at TSA.

4.5.2.3 Written simulations

4.5.2.3.1 Description

A written simulation is a kind of paper-and-pencil exercise in which a decision-making process is simulated and where the examinee has to undertake enquiries, make
decisions on the basis of those enquiries and to take suitable and appropriate action. Such written simulations entail the presentation of a realistic situation (which can be in a written, filmed, recorded, role-playing or photographic form) with all relevant information related to it and a written problem that has to be solved by the examinee. Examinees have to make choices from a variety of options and try to find an acceptable solution to the problem by working step-by-step through it (Priestley, 1982:30, 103, 107).

Priestley (1982: 30, 31, 103) states that there are certain distinctive characteristics of written simulations, namely:

- Written simulations require a series of sequential and interdependent decisions which are similar to the different steps of problem-solving (for example presentation, interpretation, analysis, resolution and so forth).
- The problem posed is normally similar to those in real situations.
- The examinees are provided with feedback after every action but are not allowed to retract their decisions on the basis of the feedback.
- Written simulations make room for a number of different solutions, but every solution must be logical and acceptable.

Written simulations are regularly used in the following areas:

- Military training (technicians)
- Survival skills
- Sales approaches
- Police investigations
- Management personnel
- Medical field
- Both for teaching and evaluation (Priestley, 1982: 31).
4.5.2.3.2 Advantages

Priestley (1982: 106, 107) mentions the following advantages of written simulations:

- They supply settings for the evaluation of problem-solving and decision-making skills that are relatively realistic.
- As in the case of real life, feedback is almost immediately available to learners after every decision that they have taken.
- Because of the availability of alternatives, examinees are free to choose whatever approach they see fit.
- Written simulations also allow learners to re-evaluate the available information and to try out different solutions to a problem.
- Written simulations are more suitable for the evaluation of the process of performance rather than products.

4.5.2.3.3 Disadvantages

Priestley (1982: 107) points out the following disadvantages:

- It takes much time and costs to develop the complex problems that are normally posed in written simulations.
- The special types of paper normally used ("opaque overlay", "latent image", X-rays, site plans, topographical maps, and so forth) are very expensive.
- Another disadvantage that the researcher can foresee, is the fact that written simulations will inevitably also suffer from most of the disadvantages of written forms of evaluation as pointed out in chapter 1, section 1.2 and chapter 3, section 3.5.1.2. This places serious limitations to its usefulness.

4.5.2.3.4 Suitability for TSA conditions

The written simulation will probably be easier to implement at TSA than actual performance measurement or simulated performance-based measurement because it
does not require equally expensive equipment. It will also be easier to administer to large numbers of learners because of the written nature thereof. However, as pointed out in section 4.5.2.3.3, written simulations still suffer from similar disadvantages as all other forms of written evaluation and will not effectively eliminate the shortcomings indicated. One of the main aims of this research is precisely to try and find alternative evaluation/measurement instruments for the written type.

4.5.2.3.5 Suitability for BM

Written simulations can be used effectively for Business Management because it will blend very well with the nature of the subject. The reasons for this assumption is the following:

- Written simulations normally pose examinees with complex problems regarding certain subject related situations or settings, which must be solved.
- Business Management as a subject lends itself for the posing of such complex problems, which is quite different to subjects in which the skills to be evaluated can be simplified easily.
- Written simulations (as mentioned in section 4.5.2.3.2) can also be administered to large numbers of examinees, which is the case with regard to BM which have student numbers running into the thousands.

However, because of its written nature, written simulations might still merely focus on the measurement of cognitive skills at the expense of actual practical and real job-related skills.

4.5.2.3.6 Financial implications

Although some of the material required for the administering of written simulations can be expensive, for example different types of pages, it is relatively cheaper than all the other forms of simulations and definitely cheaper than actual performance where expensive equipment and materials are required. In fact, written simulations would
probably not be much more expensive to administer than conventional written forms of evaluation.

4.5.2.4 Management exercises

Management exercises refer to instances in which examinees are placed into a role-playing situation and observed by examiners. Management exercises differ from simulated performance-based measurement in that it focuses on the evaluation of techniques and not on content or factual knowledge as is the case in simulated performance evaluation. A variety of management exercises will be discussed next.

4.5.2.4.1 Role-playing

a) Description

Role-playing requires from examinees to assume the role of a particular person or position in a particular company. The examinee is then presented with a certain problem which can be done in written or oral form. The examinee then has to interact with one or more persons which can be the examiners themselves, trained actors or other examinees. Examiners use checklists, rating scales or questionnaires to rate the examinee's performance and the rating is based on predetermined criteria. In many cases the examinee has to act in the capacity of a supervisor who has to solve a problem with a particular employee. The aim with such a role-playing exercise is to obtain a sample of aspects such as behaviour, interpersonal and communication skills, problem-solving skills, and so forth of examinees in a relatively realistic setting (Priestley, 1982: 111).

Rivera et al (1995:37) state that role-playing create opportunities for learners to make and justify decisions that they have made. It also helps to illustrate and evaluate the essential aspects of decision-making.
Role-playing can be used for both evaluation and training purposes. Role-playing is regularly used in the following instances:

- Candidates for administrative or management-level positions in business and industry.
- Professions that are dependent on personal interaction (for example sales and marketing).
- The military for training of officers.
- Situations involving the management of client situations (Priestley, 1982: 111, 112).

b) Advantages

- Priestley regards as the main advantage of role-playing the fact that relatively realistic settings can be provided in which a candidate's social interactive skills can be observed, yet in a controlled manner.
- He also mentions that role-playing makes it possible to evaluate skills that are not easily measured by other forms of evaluation (1982: 112).

c) Disadvantages

- Role-playing requires a great amount of time and personnel for the development, administration and scoring thereof.
- Subjectivity is probably inevitable in the rating process (Priestley, 1982: 112).

d) Suitability for TSA conditions

The tutor-system at TSA will make it possible for the arrangement of such role-playing situations. It will also be possible to conduct such sessions at TSA study centres because no great amount of equipment or instruments are required.
e) Suitability for BM

It appears as if role-playing exercises will be very suitable for BM because of the following reasons:

- It was already mentioned in section 4.5.2.4.1a that role-playing has been in regular use for the evaluation of administrative or management level positions in business or industry, and BM is obviously a similar type of subject.
- In terms of cost implications, it would seem as if role-playing is not in real terms a very expensive process to administer. The personnel requirement can easily be overcome by using well-trained peers and the time factor should also not be so severe if there are enough tutors available.

f) Financial implications

Although Priestley (1982:112) mentions the aspects of large numbers of personnel and great amounts of time that are require for the administration of role-playing tests, it would appear as if such hindrances can be overcome. As suggested in section 4.5.2.4.1e, tutors can train fellow learners and let them act in the role of required personnel. If clear instructions and guidelines are provided, it would not be much of a problem for fellow learners to perform that function.

4.5.2.4.2 Simulated interviews

a) Description

Simulated interviews are in actual fact specialised forms of role-playing but it focuses on interview skills and related behaviours. A variety of terms are used to indicate simulated interviews which include the following: “simulated diagnostic interviews”, “simulated proposed treatment interviews”, “oral simulations”, “role-playing patient management problems” and “simulated office orals”.

An example of an instance where simulated interviews are used for the purpose of certification examinations, is at the College of Family Physicians of Canada and the American Board of Orthopaedic Surgery. These interviews take place in the following way: The test situation is one in which programmed patients go for a check-up by a physician. The "patients" must then inform the physician about the "symptoms" that they are suffering from and answer any questions posed by the "physician". The performance of the "physician" is recorded on videotape and rated on the following criteria:

The ability to:
- obtain the necessary information from the patient;
- to display attitudes that are desirable and expected from a person in such a position;
- to solve the patient's problem; and
- to reduce patient anxiety (Priestley, 1982:113).

Other areas in which simulated interviews are used, are jobs in which regular interviews are required, for example:

- Journalists
- Performance appraisers
- Telephone sales
- Admissions officers
- Social workers
- Employment counsellors
- Medical and allied health fields
- Training of police officers (Priestley, 1982:113).

Simulated interviews are the same as role-playing and therefore all other facets of role-playing will apply to it.
b) Advantages

See section 4.5.2.4.1b.

c) Disadvantages

See section 4.5.2.4.1c.

d) Suitability for TSA

See section 4.5.2.4.1d.

e) Suitability for BM

See section 4.5.2.4.1e.

f) Financial implications

See section 4.5.2.4.1f.

4.5.2.4.3 Fact-finding exercises

a) Description

Fact-finding or "information seeking" exercises entail the presentation of a problem that exists within a particular company (real or fictitious). Examinees will only be provided with the minimum background information on both the problem and the company involved. Further information must be assembled by the examinee by questioning of someone that functions as a source of information. After obtaining the required information, the examinee must analyse the information and recommend a solution(s) to the problem (Priestley, 1982: 114).
The main skill that is to be evaluated is an examinee’s ability to obtain the most essential information in the shortest possible time in order to recommend a workable solution to the problem that was posed. The other related skills that are focused upon are the examinee’s ability to distinguish between relevant and irrelevant information; to grasp the essence of the problem; and to synthesise information that will lead to a solution of the problem (Priestley, 1982: 114).

Fact-finding exercises are mostly used in business and corporate settings.

b) Advantages

The following aspects can probably be regarded as the main advantages of fact-finding exercises:

- It helps to develop and promote the skills of logical reasoning and problem-solving.
- It helps learners to effectively distinguish between relevant and irrelevant, important and insignificant information.

c) Disadvantages

Priestley (1982:114) regards the following as disadvantages of fact-finding exercises:

- The fact that it is normally presented in small groups makes it possible for some learners to use the answers of other learners when their turn comes and thus reduces the opportunity for their own thinking and reasoning.
- If there is only one source of information available, it will hamper the ingenuity and individual styles of examinees, because different people use different ways to approach and solve problems.
d) Suitability for TSA conditions

Fact-finding exercises can probably be used effectively at TSA study centres because it does not require expensive material or equipment (only "resource persons" that are well versed in the subject). The small-group presentation style will also be easy to implement at the study centres.

e) Suitability for BM

It would appear as if fact-finding exercises will be an ideal form of evaluation for BM. It has already been mentioned in section 4.5.2.4.3a that fact-finding exercises are particularly used for business related fields of which BM forms an obvious part. Fact-finding exercises also blend well with the nature of BM because the problem-posing nature thereof will give opportunity for the evaluation of examinees' skills to solve the numerous problems that can arise in a business setting.

f) Financial implications

In terms of costs, it would appear as if fact-finding exercises are very economic and cost-effective forms of evaluation. Apparently no expensive instruments, equipment or materials are required and it does not require an actual job situation to be administered at.

4.5.2.4.4 Case study

a) Description

In a case study, examinees are presented with a description of a subject-related case (which can be real or fictitious). The examinee should then analyse the case and propose a solution or diagnosis of the problem. The case can be presented to examinees by means of a written document, a videotape, interview segments, medical records, and so forth. The assumption is made in the case of case studies that the
responses and behaviour of examinees in one case that serves as an example, can be
generalised to all similar cases. Priestley warns however that such generalisations
should be based on the techniques used by the examinee, rather than the contents of
the solutions, because according to him, if a technique is mastered by a student, it
would be more likely that that student will be able to apply the technique in a different
situation. Scores are based on aspects such as final solutions, recommendations, or the
process followed by examinees to reach a particular solution (Priestley, 1982: 115).

It appears as if the case study method is commonly used in business (Priestley, 1982:
115).

b) Advantages

The following aspects can most probably be regarded as some of the advantages of the
case study method:

• The examinee is normally confronted with real life type of problems which helps to
give them an idea of the actual work situation.
• It does not require great expenses to administer this type of evaluation.

c) Disadvantages

The disadvantages that can be identified by the researcher are the following:

• Unless the responses of examinees involve actual performance, the exercise will
suffer from the normal disadvantages of the written examination.
• If cases studies are not conducted in actual work settings it can become mere
theoretical exercises.

d) Suitability for TSA conditions

The same comments in section 4.5.2.4.3d will also be applicable here.
e) Suitability for BM

Refer to section 4.5.2.4.3e.

f) Financial implications

The use of case studies will most probably also not involve high expenses, especially if it is not conducted in actual work settings. However, the researcher reckons that case studies will be more effective and realistic if it is done in an actual work setting. This might cause an increase in costs, but will be more worthwhile for evaluation/measurement purposes.

4.5.2.4.5 Leaderless group discussion

a) Description

A leaderless group discussion is in fact a particular type of case-study in which a group without a leader is presented with a situation, case or problem which they must attempt to manage. In such a situation either a competition or negotiation approach, or both might be followed. The participants can either be assigned certain roles which will promote the use of projective techniques, or they might be required to act in their normal ways. Various aspects of group dynamics might be evaluated in such a way, for example:

- The ability to impose a structure upon a group
- Problem solving skills
- The skill of compromising
- The demonstration of communication skills
- The adoption of various behavioural roles (Priestley, 1982: 116).

Leaderless group discussions can be used in the training of management teams and any type of job that requires team work.
b) Advantages

The main advantages that the researcher can identify with regard to leaderless group discussions are the following:

- It facilitates the acquisition of group related skills (as mentioned in section 4.5.2.4.5a) which is a common feature of real work situations.
- The various perspectives and approaches of different learners are shared among each other and it helps individual learners to view problems from different perspectives.

c) Disadvantages

A few disadvantages of this method can probably arise, for example:

- If learners are not well-trained with regard to the principles of group discussions, their discussions will very soon go off track.
- There is sometimes the problem of too dominant participants who will not allow others a fair chance for their own inputs.
- Unless the learners get regular exposure to the real job situation, their discussions might remain theoretic because they will not take all realistic factors into account.
- It will be difficult to standardise this form of measurement, because learners cannot be evaluated individually.

d) Suitability for TSA

Leaderless group discussions can function very well at TSA study centres, provided that learners are organised into well-functioning groups. Group discussions are in fact one of the most strongly recommended approaches for distance education learners.
e) Suitability for BM

Leaderless group discussions can be a very useful form of measurement in BM, but only if it is used in conjunction with other standardised forms of measurement.

f) Financial implications

The same comments in section 4.5.2.4.4f will also be applicable here.

4.5.2.4.6 Other types

There are a few other types of simulations that will not be discussed in detail for reasons that will be mentioned below. Two examples are the following:

- Management games, which can include a great variety of possible simulations, but which are more suitable for training than for evaluation purposes (Priestley, 1982: 117) and for that reason they will not be discussed further in this chapter.

- Simulator machines which are devices that are used for the training or assessment of technical skills (Priestley, 1982: 117) and are therefore only suitable for technically orientated subjects such as motor mechanics, information technology, typing, aircraft, and so forth. For this reason it will also not be discussed any further because it will not be suitable for the nature of BM.

In conclusion to the section on simulations, the following observations can be made: The main advantage of simulations above actual performance measurement appears to be the fact that simulations allow for controlled measurement conditions. It is probably also the better option to use in instances where mistakes can have too serious consequences. It is also probably relatively cheaper than actual performance measurement. However, simulations are just as time-consuming as actual performance measurement and also involves high costs. Nevertheless, some simulations, especially
the management exercises, appear to be very suitable with regard to BM and therefore forms a very strong possible alternative to existing BM measurement instruments.

With these few comments the following category of performance-based measurement instruments will be examined next.

4.5.3 Observational Measurement

4.5.3.1 General description

Observational measurement instruments are used to score the performance of an examinee in either a controlled testing situation or in natural situations (Priestley, 1982: 34). Priestley also states that observational measurement makes it possible for the evaluator to measure performances and behaviours that are not easily or adequately measured by other kinds of tests. However he states that observational measurement is primarily subjectively inclined but that the performances and behaviours to be measured thereby, require such subjective instruments (1982:34).

Observational measurement is conducted in the following way: Evaluators observe the performance of examinees and during this process record their observations on any of the following instruments which are regarded as the main observational instruments, namely checklists, rating scales or anecdotal records. Observational measurement instruments are mostly used in combination with other measurement instruments such as work-sample tests, simulated performances, design problems and interviews (Priestley, 1982: 123).

Observational measurement instruments can be used for both product and process-based evaluation and these instruments have the following common characteristics:

- The critical components of the product or process to be measured are clearly defined.
- They all use a recording form.
The evaluator has to physically observe the product or process while focusing on the critical components indicated on the recording form.

During the course of the measurement session, evaluators must indicate on the recording form whether they have observed the critical components.

A distinction can be made between obtrusive and unobtrusive observation (Priestley, 1982: 124). Obtrusive observation takes place when the evaluator has devised a specific testing situation and requires the examinee to perform certain behaviours while observed by the evaluator. Unobtrusive evaluation on the other hand is where an evaluator observes the behaviour and performance of a candidate in a normal working situation (for example learners in a lecture hall, employees at their normal duties, and so forth) without the examinee being aware of such observation. The focus in obtrusive observation is on the observation of specific traits and behaviours that should be displayed by an examinee in a specifically assigned task. In unobtrusive observation, the observation is based on predetermined standards or criteria that are applicable to a normal situation.

According to Priestley, observational evaluation techniques can be used effectively in any area which require performance measurement, for example: business, law, medicine, arts, and so forth (1982: 125).

Woodward (1994: 23) argues that for observations to be really effective and meaningful, the recording process should be as precise or accurate as possible. To ensure such accuracy she advises that recording should be performed as close to the actual event as possible. Woodward mentions audio and video tapes as two common instruments for observation (1994: 27).

Gearheart & Gearheart (1990: 85, 86) illustrate four methods of direct observation which is based on work done by Guerin and Maier (1983). These four methods are:

- "The Chronolog": This is a form of observation in which information observed is stated in a very objective and non-judgmental way. Words that express aspects
such as evaluation, conclusions or intentions, are avoided and very short phrases or sentences are used.

- "Frequency recording": This form of observation is intended for the detection of patterns of behaviour. Such recordings can aim to reflect all the occurrences of a certain behaviour, or the number of times it occurs within a set period of time.

- "Sequence sample": Sequence samples aim at the recording of behaviour at the exact moment that a targeted problem occurs. Both the immediate consequences as well as events immediately before the occurrence of the problem are recorded. The intention is to be able to analyse behaviour in terms of stimuli, event and response.

- "Trait sample": This form of observation is based on established systems in which predetermined categories of traits exist. The observer should then record any manifestations of those traits in the behaviour of those who are being observed.

4.5.3.2 General advantages

Some general advantages of observational evaluation appears to be the following:

- The various advantages of direct measurement (validity, reliability, and so forth).
- The fact that by means of observation behaviours and performances can be measured which is not normally possible by other means (Priestley, 1982: 34).

4.5.3.3 General disadvantages

The main disadvantage of observational measurement is its subjective nature and the serious consequences it might have if incorrect or biased judgements are made. As a consequence of this subjective nature, a number of issues arose from the use of observational evaluation. The following is merely a brief summary of the main issues identified by Priestley (1982:126 - 129).
• The aspect of time-sampling: Because of the danger of incorrect judgements as a result of measurement conducted in a very short time, it became necessary to arrange for allotted time slots which will enable a representative sampling of a candidate's overall performance.

• The aspect of scoring reliability: There appears to be three factors that can influence the accuracy of the evaluators' observations and judgements, namely:

  - Personal bias which refers to the tendency of evaluators to rate all candidates at the same point on a rating scale.
  - The halo effect which refers to instances where the evaluator's general impression of a candidate might lead him to rate the candidate the same on all dimensions.
  - Logical error which occurs in cases where the evaluator over- or under-rates the level of correlation that exists between two distinct dimensions.

• The aspect of observational reports: The use of observational reports was necessitated by the fact that not all characteristics and behaviours can be observed in a brief period of time, but rather require much longer time frames.

With this brief general discussion and evaluation of observational measurement as background, a more specific evaluation of observational measurement instruments will be undertaken next.

4.5.3.4 Different observational measurement instruments

4.5.3.4.1 Checklists

a) Description

Woodward (1994: 10) states that the essential meaning of checklists is lists which are used to check that certain intended things are/were done.
A distinction is made between two main types of checklists, namely quantitative and qualitative checklists (Priestley, 1982:129). Quantitative checklists are used to evaluate the presence of particular dimensions, behaviours or attributes. No attention is paid to the quality of such an attribute. An example of a quantitative checklist is where it must be determined whether a dental hygienist follows the required steps for the cleaning of a patient’s teeth. The only focus in such a case will be to assess whether all the necessary steps were followed. No attention is paid to aspects such as speed, delicacy of movement, the quality of the conversation aimed at letting the patient relax, and so forth as would be the case in qualitative checklists. Different types of quantitative checklists can be distinguished, for example frequency checklists, participation checklists, behaviour tallies, and so forth (Priestley, 1982: 129, 130). Quantitative checklists are very useful in the observation of procedures with particular sets of steps and products with specific characteristics or components (Priestley, 1982:135).

Qualitative checklists on the other hand are used to evaluate dimensions of quality. The qualitative checklist functions on the basis of judging whether an observed behaviour or attribute is of the desired quality (Priestley, 1982:132). One requirement of qualitative checklists is that its terminology should be very clearly and precisely defined in order to prevent bias by the evaluator and to ensure uniformity. Qualitative checklists are normally required in instances where information regarding the presence or absence of quantitative dimensions is not sufficient (Priestley, 1982:135).

b) Advantages

Priestley (1982: 135, 137) lists the following advantages of checklists:

- Checklists are very useful observation tools in cases where the dimensions that will probably be exhibited during the course of observation are predictable in advance.
- Quantitative checklists have the advantage of simplicity.
- The simple yes/no responses help to reduce subjectivity by evaluators.
Qualitative checklists make possible the evaluation of the quality dimension. Validity and reliability of qualitative checklists can be increased by making use of more than one observer. Both types of checklists can be used to provide extensive feedback to examinees about their performance.

Gearheart & Gearheart (1990:86, 87) state that checklists have the advantage of being modifiable to correlate with the needs, age and unique circumstances of learners. They also state that checklists are very useful in the cases of planned interventions or remediation of learners.

c) Disadvantages

The following disadvantages of checklists are identified by Priestley (1982: 135, 137):

- Checklists cannot be of much use in situations where the anticipated behaviours are unpredictable.
- Quantitative checklists are basically a crude measurement instrument because it does not include finer details of what is to be observed.
- Qualitative checklists are sometimes too rigid and inflexible with regard to steps or procedures.
- Qualitative checklists can suffer from the problem of high subjectivity.
- Both quantitative and qualitative checklists are quite expensive and time-consuming in terms of development, administering and scoring.

Woodward (1994: 10) regards as the greatest disadvantage of checklists the fact that it cannot adequately portray individual progress.

d) Suitability for TSA conditions

It might be possible to use checklists at TSA's study centres by tutors for the measurement of student progress, but only to a limited extent. The reason for this
contention is the fact that tutors are not appointed on a full-time basis and might not be available for such time-consuming measurement processes such as checklists. Checklists, as a type of observational measurement instrument apparently requires a lot of time and contact between evaluator and student. Tutors will also need proper training in the use of checklists.

e) Suitability for BM

Checklists can be used in BM for the recording of certain subject-related skills. However it will not provide adequate records of students' progress and achievements and can therefore not be used independently as a measurement instrument. The time-consuming nature of checklists and the requirement of regular contact between evaluator and student for observational purposes, might hamper the usefulness of this instrument in the context of TSA.

f) Financial implications

Checklists, although it also involves certain costs, might not be that expensive to use, compared to other types of performance-based measurement instruments. However, its time-consuming nature might contribute towards expenses regarding its implementation and use.

4.5.3.4.2 Rating scales

a) Description

Rating scales operate in the same way that checklists do, with the only difference that instead of merely a yes or no mark next to a dimension, the degree or extent to which a particular dimension is present, is indicated in the case of rating scales. The rating takes place on a continuum which can range from yes to no, positive to negative, and so forth. Some examples of such continuums are the following:
• Often, sometimes, seldom, never

• Extremely well-unified, coherent, clearly articulated, harmonious (Priestley, 1982: 137, 138).

Priestley (1982: 138 - 143) distinguishes four types of rating scales, namely:

• Numerical rating scales are scales in which the ratings appear in the form of numbers.

• Graphic rating scales which use words as ratings.

• Descriptive-graphic rating scales in which a detailed description of each point of the graphic scale is provided.

• Ranking scales in which products or processes are rated in comparison to products and processes that are similar to it.

Rating scales are mostly used for the indication of affective traits such as attitudes and opinions because such affective traits can only be measured on a continuum and not on a right/wrong basis.

b) Advantages

The following advantages of rating scales can be distinguished:

• Numerical rating scales are very easy to use because when compiling total scores the numbers merely have to be added.

• Graphic scales are more flexible and self-explanatory than numerical scales.

• Descriptive-graphic rating scales, if used in appropriate settings, can be very precise and as a result of such precision, reduces subjectivity to a great extent.

• Ranking scales (especially the criterion-referenced ranking scale) can be used by different evaluators to evaluate various products or performances in relation to uniform standards.
c) Disadvantages

The following disadvantages of rating scales can be identified:

- Numerical rating scales cannot effectively measure behaviours in cases where there is a changing set of criteria because each number must always have the same meaning for every dimension that is listed.
- In the case of graphic ratings, the one- or two-word descriptions are not as precise as it might seem and it paves the way for great subjectivity (for example, different people might interpret the words differently).
- It is very difficult to convert graphic ratings to numbers if required for scoring purposes.
- Descriptive-graphic ratings are unable to provide precise feedback to a candidate because of the various aspects in every description.

d) Suitability for TSA conditions

Refer to section 4.5.3.4.1d.

e) Suitability for BM

Refer to section 4.5.3.4.1e.

f) Financial implications

Refer to section 4.5.3.4.1f.

4.5.3.4.3 Anecdotal records

a) Description

Anecdotal records comprise written descriptions of important events in the life or work of various people. Such records are normally compiled by educators and supervisors
about their learners or employees. Anecdotal records serve to capture events that are normally forgotten or which might not be remembered correctly. These records can make it possible for educators and supervisors to detect significant patterns of behaviour in the lives of learners or employees which can assist in the making of diagnoses or recommendations. Anecdotal records are not used for the purpose of judging, but only to describe certain important events (Priestley, 1982: 149).

Woodward (1994: 11) describes anecdotal records as a very popular recording instrument among learners. However, she mentions that, probably because of a lack of experience, educators in her research project did not make real use of such anecdotal records of learners and continued with prior planned programmes that ignored the contents of such records.

Anecdotal records are normally used in the following instances:

- To describe events involving social adjustment.
- To describe natural behaviour.
- To describe individual idiosyncrasies.
- It can help to identify individuals who need special attention.

b) Advantages

The following advantages of anecdotal records can be mentioned:

- A main advantage of anecdotal records is the fact that it provides a more extensive description of actual behaviour than rating scales or checklists.
- It makes possible the recording of all significant events including atypical ones.
- Anecdotal records also make possible the recording of unexpected and unusual events and this aspect makes it a very useful tool in instruction programmes.
- Because of its less formal nature, it is very suitable for performance measurement which requires such informal evaluation (Priestley, 1982: 151, 152).
c) Disadvantages

Priestley does not mention any disadvantages of anecdotal records, but based on a logical scrutiny of anecdotal records, a few disadvantages can be identified, namely:

- Anecdotal records as a form of evaluation cannot be standardised.
- Since anecdotal records are not used for the purposes of scoring or judgement, it cannot function as an independent form of evaluation.
- Anecdotal records are not based on prior set criteria.

Another possible disadvantage pointed out by Woodward (1994: 11) is the fact that it appears in practice as if educators seem to concentrate on the recording of learners that gain their attention because of difficulties they seem to experience. Such records are normally negative and the more quiet type of student is basically neglected.

d) Suitability for TSA conditions

Anecdotal records suffer from the same disadvantage of checklists, namely its time-consuming nature. However, it can provide educators with useful information regarding learners which is normally overlooked. It can also help TSA educators to gain a better understanding of learners and the backgrounds they come from.

e) Suitability for BM

Anecdotal records can probably be used as an additional or complementary instrument to provide lecturers with more personal information regarding learners. However, because of the fact that it cannot be easily standardised, it cannot function as a major measurement instrument in BM.

f) Financial implications

Refer to section 4.5.3.4.1f.
In conclusion to the section on observational measurement, it can be justly stated that observational measurement goes hand in hand with actual performance measurement as well as simulations, because as in the case of these two types of measurement, observational measurement also involves the measurement of the actual process of behaviour and demonstration of skills and not merely a completed product. As a direct form of measurement it is also probably one of the most ideal ways to measure student performance and skills, however, its time-consuming nature makes it a very difficult instrument to implement in distance education. Even though TSA’s tutor system might alleviate this problem, the part-time functioning of tutors might still make it difficult to implement effectively. However, if ways to overcome this obstacle can be devised, observational evaluation will be a very suitable form of measurement for TSA learners and for BM learners in particular.

With these closing comments, a further dimension of performance-based measurement will be examined next.

4.5.4 Oral Measurement

Although oral measurement is one of the oldest forms of measurement (see chapter 2 section 2.2.1.1), its modern formats, for example interviews, presentations, and so forth, can indeed be regarded as alternatives to written examinations. Oral measurement techniques are based on oral responses. Priestley (1982:153) states that oral measurement can be used for any performance-based test provided that it does not require physical actions (for example repairing a motor engine). He emphasises the fact that oral measurement requires a lot of practice in order to perfect it.

Oral measurement is very appropriate in the following types of situations:

- In cases where examinees have to defend themselves or explain their own reasoning.
- In instances where oral communication skills form the essence of performance.
• In instances where it is impractical or impossible to use written evaluation techniques (Priestley, 1982: 153).

Instruments used for the recording and scoring of oral measurement are normally checklists or rating scales. Oral measurement is normally administered on an individual basis or in small groups. The scoring normally takes place while the test is administered.

The main advantage of oral measurement is the fact it makes possible the evaluation of skills or behaviours that cannot be evaluated by means of another evaluation method. However, there are some disadvantages involved in oral measurement which include aspects such as:
• High time and cost implications.
• Difficulty in obtaining valid and reliable scoring, and so forth (Priestley, 1982: 154).

The rest of the advantages and disadvantages of oral measurement will be illustrated during an examination of specific forms of oral measurement. The three main instruments that will be examined are:

• Oral examinations
• Interviews
• Prepared presentations

4.5.4.1 Oral examinations

4.5.4.1.1 Description

Oral examinations are not a new phenomenon. In fact it has a very long history. Swanson, Norman & Linn (1995: 6) for example, mention that oral examinations have been used for a few hundred years in countries under the British Commonwealth in the past to measure clinical skills for medical practice. They even mention that although
oral examinations were removed from US licensing examinations, most of the US and Canadian medical speciality certification examinations still use oral examinations.

There are apparently two forms of oral examinations, namely:

- The oral tests/quiz which is used to measure more general knowledge contents.
- The oral defence which is normally used to measure more specialised knowledge domains (Priestley, 1982:154).

Questions in the oral test or quiz normally focus on essential knowledge content which is required by examinees to be able to perform effectively in a particular area of competence. These questions are normally open-ended. Answers by examinees can be very brief or as long as 10 to 15 minutes.

In the case of an oral defence, the questions are based on a particular product developed by the examinee. Such a product can be for example a dissertation, thesis, design project, and so forth. The examinee is then required to defend this product before two or more examiners.

Oral examinations can be structured and formal or unstructured and informal.

A few examples of areas or institutions where oral examinations are regularly conducted are the following:

- The California Board of Architectural Examiners
- The American Academy of Orthopaedic Surgeons
- The College of Family Physicians of Canada
- The "oral trade test"
- In education
- In the evaluation of foreign language proficiency
- Speaking ability in various languages
- In tertiary education (the oral defence) (Priestley, 1982: 155 - 160)
4.5.4.1.2 Advantages

Priestley identifies the following advantages of the oral examination:

- Oral examinations make it possible to evaluate certain complex skills that cannot be adequately measured by other measurement instruments. These skills include aspects such as:
  - oral communication skills
  - "fast thinking" skills
  - the ability to apply relevant knowledge to a problem at hand.
- Oral examinations grant examinees the opportunity to justify or explain their responses.
- Oral examinations allow for follow-up questions which help to clear up misunderstandings (This is a very important aspect because in the case of written exams no such opportunity exists for examinees and they might fail a paper because of a mere misunderstanding).
- The face validity of oral defences is relatively high because it resembles a common practice in the corporate world (1982: 160, 161).

4.5.4.1.3 Disadvantages

Priestley also identifies the following disadvantages of oral examinations:

The main disadvantage of oral examinations is the fact that it can normally only be conducted on one individual at a time. This factor causes the following disadvantages:
- It becomes a very time-consuming process for both examiner and examinee
- It becomes a very costly practice.
- The other main disadvantage is the aspect of subjectivity which in turn has other negative consequences, for example:
If rating forms are used they might also suffer from shortcomings such as inconsistencies among raters, disagreement concerning assessment standards, and so forth.

The "halo-effect" (instances where the first impressions of a rater can influence all further scoring/rating by that rater) might be at work (1982: 161).

4.5.4.1.4 Suitability for TSA conditions

Because of the large numbers of learners at TSA, the oral examination might not be the most feasible type of measurement, because of its time-consuming nature. Even in the tutor system, the time aspect will not be adequately addressed. However, on a formative or continuous measurement basis oral exams can be used by tutors.

4.5.4.1.5 Suitability for BM

With regard to BM, oral examinations can play a useful role. It will allow learners to practice and prove their verbal skills with regard to the subject. Especially a learner's communication skills within the job context can be tested provided that real settings are used.

4.5.4.1.6 Financial implications

It has already been mentioned in section 4.5.4.1.3 that oral examinations can be an expensive process. The time-consuming nature of oral measurement also contributes to the costs aspect.

4.5.4.2 Interviews

4.5.4.2.1 Description

With regards to measurement, an interview is a situation in which the knowledge or skills of examinees are evaluated by examiners on the basis of predetermined criteria.
This evaluation is conducted in a dialogue form. The questions asked by the examiner are aimed at obtaining a better understanding of the examinee. The questions can be strictly factual or affective or vary between the two extremes. The questions are intended to establish the qualifications or skills of a candidate with regard to a particular training programme, a job, a degree, a license, or a certificate. The questions might focus on aspects such as knowledge, skills, experience or attitudes (Priestley, 1982:165).

The two most common uses of interviews as a measurement instrument are for screening and for selection purposes. Selection interviews are normally used to determine whether a candidate should be accepted into a certain course of study or a job, and so forth. In such cases the focus is on a candidate's probability of success in that course of study or whether they match a particular job. In the case of screening interviews, the interview is used to determine the eligibility of candidates to receive a particular license or certificate. In such cases the focus falls on aspects such as attitude, character, potential to make a contribution to the field, and so forth (Priestley, 1982:166).

Gearheart & Gearheart (1990: 90, 91) state that the value of interviews is situated in the accuracy of the information obtained from a particular person(s). They state the need for well-trained interviewers. Interviews also allow educators to see certain aspects from another person's point of view. It also provides educators with valuable information regarding students' behaviour and so forth.

A number of common attributes across subject fields have been established. These include aspects such as:

- Content expertise directly related to a particular situation
- Training
- Work experience
- Oral communication skills
- The ability to address professional problems
Team-work skills
Safety-consciousness
Levels of professional responsibility and accountability
Confidence and professional conduct
Interest in the career and upward mobility (Priestley, 1982: 166, 167).

Priestley recommends that the questions in an interview should be constructed in a way that will elicit responses that are:
- factual
- verifiable
- specific and

Interviews are regularly used in the following fields:

- Education (early childhood and elementary school)
- As a screening device (for example to eliminate unqualified candidates)
- As a student evaluation technique
- As a programme evaluation instrument
- In professions such as:
  - physicians
  - counsellors
  - psychiatrists
  - social workers
  - police officers

4.5.4.2.2 Advantages

The following advantages of interviews as a measurement instrument can be distinguished:
• The interview is a very individualistic and personalised form of evaluation and thus allow for comprehensive measurement.
• It makes possible the evaluation of certain essential personal characteristics which are crucial for good performance in a job or educational setting.
• Interviews allow the examinee the opportunity to obtain more clarity from the examiners about what they are supposed to do and thus helps to eliminate misunderstandings.
• Examinee questions and responses make it easy for examiners to establish the extent of command they have in a particular field.
• Interviews are very flexible and can be tuned in to the particular level of any individual examinee (Priestley, 1982:170).

4.5.4.2.3 Disadvantages

The following disadvantages of interviews have been identified:

• Interviews do not provide information that can be standardised for formal evaluation and is therefore more useful as an informal measurement instrument.
• Various sources of mistakes can occur in the rating process, for example:
  - disagreement concerning the standards of competence among examiners;
  - incorrect interpretation of evaluation forms by examiners;
  - the halo-effect;
  - low consistency among interviewers.
• Interviews can only measure a limited number of human characteristics.
• Interviews suffer from low accuracy with regard to predictions of future success by examinees.
• It can also be a very time-consuming and costly exercise.
• Interviews are more effective as a training or instruction instrument than a measurement instrument (Priestly, 1982:170, 171).
4.5.4.2.4 Suitability for TSA conditions

As a result of some of its disadvantages such as the difficulties concerning standardisation, its time-consuming nature, its low predictability value, and so forth, interviews will only be suitable as an informal type of measurement instrument which can help to complement other, more formal types of measurement instruments.

4.5.4.2.5 Suitability for BM

The same comments in section 4.5.4.2.4 will also apply in this case.

4.5.4.2.6 Financial implications

Because of its individualistic and time-consuming nature, interviews can probably become a very expensive form of measurement if used on large scale.

4.5.4.3 Prepared presentation

4.5.4.3.1 Description

Prepared presentations involve the organised display of certain specified information by examinees. It is normally conducted in the format of a lecture or speech, but can also include the use of audio or visual aids. Prepared presentations can be used for the following evaluation purposes:

- To evaluate a certain product.
- To determine the extent of knowledge a particular candidate has in a particular field.
- To evaluate skills such as oral communication, information selection and organisation, and so forth (Priestley, 1982: 175).
When assigning examinees the task of a prepared presentation, certain parameters are set within which such a presentation should fall. Such parameters include:

- Time (that is the length of administration)
- Content
- Preparation time
- Costs to be incurred (where applicable) (Priestley, 1982: 175).

The following options can also be followed in the assigning of prepared presentations:

- All examinees can be given the same assignment.
- A limited choice of topics can be allowed; or
- An unlimited choice in topics and freedom in style of presentation can be allowed.

For scoring purposes, standard rating forms can be used. Such standard rating forms are very useful in cases where there are a number of examiners involved in the rating process.

A few examples of areas or situations in which prepared presentations are normally used are the following:

- When applying for certain types of jobs (especially academic or professional).
- The dramatic arts.
- Areas that require public speaking (for example politicians).
- Teacher trainees.
- The demonstration of a particular product (for example acting skills in a film) (Priestley, 1982: 176, 177).

4.5.4.3.2 Advantages

The following advantages of prepared presentations can be mentioned:
They normally encompass most of the advantages of the previous two forms of oral measurements (see section 4.5.4.1.2 and 4.5.4.2.2).

They are effective instruments to measure certain oral communication skills that might form part of certain jobs.

Because of the fact that a large number of examiners are normally involved, very reliable ratings can be produced (Priestley, 1982: 177).

4.5.4.3.3 Disadvantages

Disadvantages of prepared presentations are the following:

- It can be a costly process because of its individualistic nature (in terms of examinees).
- It involves lengthy administration and preparation.
- A large number of examiners can lead to increased costs, time and scheduling problems.
- If rating forms are used, the same disadvantages will apply as mentioned in section 4.5.3.4.2.3.
- In the case of an absence of rating forms and a small number of examiners, there is a threat of high subjectivity.
- Where there is only a small number of examiners involved, there is a tendency to consider personal characteristics that are unrelated to the evaluation concerns (Priestley, 1982:177).

4.5.4.3.4 Suitability for TSA conditions

Prepared presentations, because of its individualistic nature and the fact that it requires a number of examiners at the same time to make it more objective and reliable, will probably limit its effectiveness for TSA with its large student numbers.
4.5.4.3.5 Suitability for BM

Although certain content areas in BM might warrant the use of prepared presentations, it does not form a very dominant feature of the subject's nature.

4.5.4.3.6 Financial implications

Refer to section 4.5.4.2.6.

In conclusion it can be said that oral measurement can also fulfil a very useful role, because it is the ideal instrument to use in instances where communication-related skills are being measured. It can be used in conjunction with other measurement instruments and despite its time-consuming and individualistic nature, it should be strongly considered for greater use at TSA.

With these brief comments the next category of performance-based measurement will be addressed next.

4.5.5 Programme Requirements

Programme requirements is a term used by Priestley (1982:219) and which in essence refers to "non-tests" requirements, or differently put, those measurement-related practices that are not formal tests, but which can have an influence on the formal evaluation process. Examples of such programme requirements are resumes, portfolios, transcripts, references, proof of qualifications and comparisons of student performances with those of other learners. Such records of student performance can function in conjunction with the formal evaluation process.

Priestley (1982:219) however states that one shortcoming of programme requirements is the fact that it is generally based on assumptions and may therefore in many cases not be able to yield reliable results. As a result of this factor he advises that programme requirements should be used with great care in the evaluation process. He
also mentions that in the case of employment or licensing and certification programmes, the people in charge will never use programme requirements as the determining factor in selecting a candidate. Some of the reasons for this almost scepticism about the meaningful use of programme requirements stem from disadvantages such as:

- The fact that it is an unstructured process.
- The potential bias that might be present in the process.
- The highly subjective nature thereof.
- The fact that they are very prone to subjective and arbitrary uses.
- They can rarely be regarded as proof of actual on-the-job competency because they mostly record past performance (Priestley, 1982: 219, 220).

Programme requirements however also entail a number of important advantages which include the following:

- They can provide very valuable information for evaluation purposes.
- They can function well in conjunction with other forms of formal evaluation.
- In some cases programme requirements are the only practical method of obtaining certain information about candidates (Priestley, 1982: 220).

The three main categories of programme requirements which will be examined in this chapter are: personal records, performance records and apprenticeship or internship.

4.5.5.1 Personal records

4.5.5.1.1 Description

In the glossary of Broadfoot (1986: 237) personal records are defined as: "A record initiated by the pupil describing experiences and achievements considered of value and significance to that pupil".
The purpose of personal records are mainly to portray the aptitude or competencies of a certain candidate for a job, profession or study course. These records include information on aspects such as a person's:

- Personal background.
- Education.
- Previous experience.
- The perceptions of others about that person.
- Any form of training that the person underwent (Priestley, 1982:220, 221).

Personal records can be used both for selection and screening purposes. Quite a variety of personal records exist, for example: Records for the indication of potential ability which consists of:

- Academic transcripts which are verified records provided by schools or special programmes in which the academic courses and grades that learners received, are recorded.
- Recommendations or references which are written or verbal descriptions of learners by their supervisors, advisors, instructors or colleagues.
- Personal statements which are written essays in which candidates write about their own strengths and aspirations with regard to a particular programme or job.
- Apprenticeship records or job history which are records supplied by candidates in which they give an account of previous training and job experiences that they have obtained (Priestley, 1982:221, 222).

A few examples of purposes and areas in which personal records are used are the following:

- For making decisions about:
  - college entrance
  - graduate school placement
  - promotion awards
- certification
- eligibility to take certification examinations

- In architecture.
- The Association for Education for the Visually Handicapped.
- The American Association for Musical Therapy.
- The Continuing Medical Education programmes (Priestley, 1982: 222, 223).

4.5.5.1.2 Advantages

The following advantages of personal records can be identified:

- It provides important background information about candidates that can contribute towards predictions of future performance.
- In many cases personal records complement each other in terms of information and sources.
- It is a relatively cost-effective and time-saving form of measurement (Priestley, 1982: 223, 224).

4.5.5.1.3 Disadvantages

The following disadvantages of personal records can be pointed out:

- They suffer in terms of completeness and accuracy.
- It is sometimes difficult to compare the records produced by different people, programmes or institutions because there is no uniformity in terms of grades and standards.
- In the case of personal records, there is a tendency among candidates to say what they think examiners expect from them and that might not be a true reflection of themselves.
- Personal records cannot be used as the sole source of information for evaluation purposes (Priestley, 1982: 224).
4.5.5.1.4 Suitability for TSA conditions

Personal records can fulfil a very useful complementary role in terms of measurement of students' knowledge, skills and abilities. Because of one of the main shortcomings of distance education, the absence of face to face contact between lecturer and student, learners remain merely a number to lecturers at the main campus. Personal records can help to overcome this problem and can serve to provide educators at TSA with more background information about learners which can be taken into account in the final evaluation process. Personal records can also provide useful information to educators which can help to enhance motivation strategies used by them.

4.5.5.1.5 Suitability for BM

Personnel records can also help to supply additional information about learners studying BM which can enable lecturers and tutors in BM to draw links between subject contents and students' environments or sources of reference. This might contribute to student performance and achievements in the subject.

4.5.5.1.6 Financial implications

The financial implications of personal records will be minimal to TSA, because the full process will be conducted by the learners themselves.

4.5.5.2 Performance records

4.5.5.2.1 Description

Performance records are slightly different from personal records because the emphasis in performance records falls on an indication of the quality or level of performance in every qualification or certificate obtained by a particular candidate. In the case of personal records the emphasis falls on a mere indication of particular requirements that a candidate complies with. Where personal records are quantitative in nature,
performance records are qualitative. Performance records are therefore a form of evaluation report that indicates the actual performance that might be required by a particular programme (Priestley, 1982: 224).

A few examples of performance records are:

- **Writing samples:** Writing samples entail mostly written descriptions of work-samples. Writing samples commonly serve as required submissions in professions where writing proficiency is a required skill (for example journalism, copy writing, technical writing, and so forth) (Priestley, 1982: 225).

- **Resumes:** Resumes are used as a screening device to help identify candidates from whom more information might be required, and thus makes possible a qualitative evaluation of the performance records of candidates (Priestley, 1982: 225).

- **Personnel files:** Personal files are regularly used in businesses to record important information about employees which can include aspects such as their training records, production records, awards obtained, salary increases, promotions if any, reprimands, and so forth (Priestley, 1982: 227).

- **Progress reports:** Progress reports are mainly used to provide information about the training and education records of employees. They are regularly used in professions such as business, training and education (Priestley, 1982: 227).

- **Medical records:** Medical records are a specialised type of assessment in which the performance of medical personnel (for example doctors, physicians, and so forth) is evaluated on the basis of medical records of patients treated by them. They are normally only used in the medical profession.

- **Checklists and rating scales** are also commonly used in the evaluation of these different types of performance records.
• The most commonly and widely used types of performance records are portfolios and profiles or records of achievement. Because of the dominance of these two types they will be discussed in more detail.

4.5.5.2.2 Portfolios

Priestley (1982: 225, 228) describes portfolios as a collection of visual samples (for example drawings, sketches, pictures, paintings, sculptures, buildings, and so forth) of the actual work or previous performance of a candidate. Portfolios, according to Priestley serve to provide samples of the work done by a candidate which can in turn be used to judge candidates' level of mastery of skills and the amount of experience they have obtained. Priestley states that portfolios are used in most professions for the purpose of employee selection.

Gordon & Bonilla-Bowman (1996: 36, 37) describe portfolios as a method of record-keeping. Its function is normally to record and systematise information concerning aspects such as:
• students' teaching and learning experiences;
• the products of such teaching and learning experiences;
• samples of previous performances;
• the reflection upon these sources of information by learners, their parents and educators.

Johnsen (1996: 12) describes portfolios as "collections of work" from learners which can include both performances and other forms of work.

Fisher & King (1995: 4) define portfolios as: "a multifaceted assignment that expects more than one type of activity and production for completion and that extends over a period of time." They distinguish between the following types of portfolios:

• "The working portfolio": In this type of portfolio the evaluation process is conducted by educators and learners jointly.
• "The showcase portfolio": In this type of portfolio the emphasis falls on a display of the most outstanding work of the particular learners.

• "The record-keeping portfolio": This type of portfolio usually contains those performances and work of learners that are not included in the showcase portfolio, but which is necessary for other recording purposes (1995: 12).

Gordon & Bonilla-Bowman (1996: 37, 38) mention the following aspects about portfolios which can be regarded as some of its advantages:

• Portfolios can serve as an instrument for the development of both teaching and learning.
• It captures information about educational processes and outcomes.
• It displays cognitive and affective elements of teaching.
• It helps individual learners to internalise criteria and to establish their own personal standards.
• Portfolios allow learners to reflect on the work that they have done as well as the processes that were involved.

The use of portfolios also has the following consequences for learners:

• It helps to create a metacognitive awareness in learners.
• It encourages learners to reflect on their work and to revise it regularly.
• Portfolios enable learners to use different modalities of learning.
• Portfolios help teachers to understand their learners better.
• Learners experience the evaluation process more as a form of instruction.
• It helps learners to become aware of the ties between home and the place of instruction.
• Learners are better able to recognise standards and criteria and to understand its importance in the learning process (Gordon & Bonilla-Bowman, 1996: 39 - 41).

4.5.5.2.3 Records of achievement/profiles

Assiter & Shaw (1993:20) mention another type of performance record, namely
records of achievement or profiles. According to them these terms are used alternatively by different authors, but the same meaning is attached to both terms. This view is also held by Mansell (1986: 24). He states that in incidences where a distinction is drawn between the two terms (records of achievement and profiles) there appears to be some confusion about what each actually entails. One viewpoint sees a profile as a smaller part of a record of achievement, while another viewpoint sees it just the opposite way.

The essential and common elements of both terms are documents that record students' evaluation results and any other information related to their educational progress (Broadfoot, 1986: 238). Some basic examples of the exact instruments used for records of achievement/profiles are logbooks and diaries (Mansell, 1986: 25).

Hall (1989: 16) claims that records of achievement can be used as a summative record for school leaving purposes. He identifies the following advantages of records of achievement:

- It allows considerable inputs from learners and creates a type of partnership between learners and their educators.
- Because of their involvement learners feel more valued and it creates in them a very positive attitude toward their studies and the evaluation process.
- Students' strengths and weaknesses can be identified more easily and be addressed.
- The information obtained by means of records of achievement can serve to facilitate modifications in teaching styles and objectives where necessary (1989: 21, 22).

Hall (1989: 33, 34) however also identifies a few shortcomings of records of achievement. These shortcomings are:

- The fact that the recording process can be a very detailed and demanding exercise.
- The process of recording achievement has not yet been standardised.
Some central characteristics of records of achievement/profiles mentioned by Assiter & Shaw are the following:

- It is a process that helps to empower learners to active involvement in the evaluation and monitoring of their own development and learning.
- It involves a process of dialogue between student and educator which is characterised by student self-reflection, giving feedback, identifying own strengths and weaknesses and determining own learning aims and objectives.
- It is a process that is primarily formatively oriented, but which allow room for summative reports (1993: 20, 21).

Assiter & Fenwick (1993:26) distinguish between various types of profiles. The three main types identified by them are:

- The prescribed learning outcomes profile: This type of profile consist of a description of a student's learning outcomes for a course or part of a course.
- The negotiated learning outcomes profile: In this type of profile the student can not only identify and describe their own learning outcomes, but can negotiate concerning the following aspects: the approach by which the desired outcomes can be obtained; the type of measurement instruments to be used; and what is to be regarded as adequate proof of achievement of learning outcomes.
- The personal development profile: This type of profile focuses on formative development of learners and self-evaluation forms a major component thereof.

4.5.5.2.4 General advantages of performance records

The following general advantages of performance records can be identified:

- Performance records can effectively demonstrate the skills acquired by applicants and their abilities to perform specified tasks.
- It (for example portfolios) allows for candidates' own choice with regard to the inclusion of past experiences or jobs performed in such records.
• It also makes it possible for candidates to present the best image of themselves (Priestley, 1982:229).

4.5.5.2.3 General disadvantages of performance records

According to Priestley (1982: 229) most disadvantages of performance records stem from difficulties with regard to standardisation. These disadvantages include the following:

• The inherent subjective nature of performance records.
• Performance records are not in actual fact designed for the measurement of large quantities of factual knowledge.
• Performance records cannot in fact provide complete records with regard to professional positions of candidates.
• Performance records are in fact unable to provide accounts of the actual processes involved in certain jobs or the production of a certain product, because they are normally “after-the-fact” records.
• Performance records can also not serve as sole forms of measurement and must be used in conjunction with other measurement instruments.
• Certain types of performance records (for example personal files and progress reports) deny candidates the opportunity to render direct inputs.

4.5.5.2.4 Suitability for TSA conditions

Performance records, similar to personal records, can provide lecturers at TSA with valuable background information about their learners. Its qualitative nature can also provide lecturers with information regarding a student’s level of cognitive and other development and skills. Most types of performance records allow learners to participate in the measurement/evaluation process because it is based on their own inputs and views about themselves. This in turn contributes a lot towards the motivation levels of learners. However, similar to personal records, they can only be used in a complementary way.
4.5.5.2.5 Suitability for BM

The qualitative nature of performance records with regard to level of performance or achievement, can also assist BM lecturers in the design and development of their study material because they will have a better understanding regarding the familiarity of learners with regard to the subject. As mentioned in the case of personal records (section 4.5.5.1.5), performance records can also help to guide lecturers with regard to the most suitable instructional styles or approaches to follow.

4.5.5.2.6 Financial implications

Refer to section 4.5.5.1.6.

4.5.5.3 Apprenticeship/Internship

4.5.5.3.1 Description

An apprenticeship or internship differs from personal and performance records because the first two has to do with the evaluation of products, while apprenticeship/internship has to do with the actual processes involved. Apprenticeship/internship therefore involves direct observation of the actual performance of candidates. It can however also include product evaluation. Apprenticeship/internship can also be regarded as programme requirements because in certain subjects or fields it may be an essential part of the total educational or training programme without which no certification or licensing can occur (Priestley, 1982:230).

Apprenticeship/internship programmes are supposed to be well-structured and closely monitored in order to make it possible for candidates to learn step by step from direct and practical “on-the-job” experience. However, this is rarely the case because in most companies or industries where apprenticeship/internship programmes exist, candidates
are thrown in at the deep end and have to struggle by themselves to “swim” (Priestley, 1982: 230).

As a measurement instrument, apprenticeship/internship programmes allow for direct on the job observation of the application and mastery of actual skills. Various techniques can be used to record the information, for example work-samples, supervisor ratings, observational assessment techniques, personal and performance records (Priestley, 1982: 230).

Apprenticeship/internship programmes are mostly used in fields where a lot of practical or technical skills are required, and in which mistakes on the job can have very serious consequences. Examples of such fields are:

- **Trades** such as:
  - Carpentry
  - Plumbing
  - Boilermaking

- **Vocational areas** such as:
  - Journalism
  - Advertising
  - Banking

- **Internships** in such fields as:
  - Communication
  - Teaching
  - Medicine
  - Allied health
  - Architecture

4.5.5.3.2 Advantages

- The main advantage of apprenticeship/internship is the fact that it involves direct observation of the actual application of skills and competencies.
• It also helps to ensure that future on-the-job mistakes are prevented or minimised because licensing or certification will not take place unless a candidate has mastered the required skills (Priestley, 1982:233).

• A very important advantage that the researcher can foresee is the fact that apprenticeship/internship programmes will eliminate any possibility that learners will receive a particular qualification certificate without having mastered the actual real job-related skills. (This is a main criticism against tertiary educational institutions, including TSA).

4.5.5.3.3 Disadvantages

The main disadvantage of apprenticeship/internship programmes is the difficulties that are experienced with regard to standardisation. Priestley (1982: 233) states that it takes enormous efforts to try and standardise the processes involved.

4.5.5.3.4 Suitability for TSA conditions

Apprenticeships/internships or experiential learning as it is known at TSA, is probably one of the most ideal forms of evaluation/measurement for TSA. TSA is faced with the dilemma that, because of its distance teaching and part-time learning nature (correspondence), it initially aimed its courses mostly at learners that were employed. However, TSA's learner population has changed dramatically over the past few years and there is currently a large number of unemployed and in fact full-time learners. Especially for this type of learner the introduction of apprenticeship/internship programmes in industry and other places of possible employment will be most beneficial. The practical and skills driven focus of TSA as a technikon actually makes the implementation of apprenticeship/internship programmes in all disciplines a must (at least on third year level).
4.5.5.3.5 Suitability for BM

Apprenticeships/internships will be very suitable to BM learners, especially on third year level, because it will prepare and equip them with a proper understanding of the skills and duties that will be required from them in industry.

4.5.5.3.6 Financial implications

The implementation of apprenticeship/internship programmes might be an extremely costly process. However, if industry and other relevant bodies can be involved in the total education process on a voluntary and supportive basis, such programmes can become a reality very soon. There is already an existing level of liaison between TSA programme-groups (departments) and industry at large. This liaison should be deepened and the possibility of apprenticeship/internship programmes should be high on the agenda.

4.5.5.4 Questionnaires

4.5.5.4.1 Description

Although questionnaires are not mentioned as a type of programme requirement, it was placed under this category because it is a very commonly used instrument to obtain personal information from individuals. Much of the information in personal and performance records are in many instances obtained by means of questionnaires.

Gearheart & Gearheart (1990: 91) state that questionnaires can almost be regarded as written interviews. However, they state that the questionnaire is subject to a few limitations compared to interviews.
4.5.5.4.2 Advantages

Gearheart & Gearheart (1990: 91) mention a few advantages of questionnaires, namely:

- Questionnaires are a more permanent form of data recording (compared to interviews) and can be used for a long period of time.
- Questionnaires also allow for a more relaxed atmosphere for the student compared to interviews where learners feel threatened by the presence of the interviewers.
- A further advantage foreseen by the researcher is that questionnaires can be used by distance education educators to obtain personal information about learners.

4.5.5.4.3 Disadvantages

Gearheart & Gearheart (1990: 91) mention the following shortcomings or disadvantages of questionnaires:

- There is a danger of misinterpretation of questions.
- Respondents are limited by the fixed format in terms of expressing themselves.

4.5.5.4.4 Suitability for TSA-conditions

Questionnaires can serve a useful function at TSA in providing the opportunity to lecturers to get some more personal information about their learners which can include information about the student as a person, the environment learners live in, their socio-economic status, the main norms and traditions that learners are mainly exposed to and so forth. However it will also merely fulfil a complementary function to other forms of measurement.

4.5.5.4.5 Suitability for BM

Refer to section 4.5.5.4.4.
4.5.5.4.6 Financial implications

The use of questionnaires will probably not cause great expenses to TSA compared to the other forms of performance-based measurement. The only real costs will be with regard to paper and postage.

In conclusion it would appear as if especially personal records and performance records can provide evaluators with very useful and essential information regarding their learners. The main advantage thereof is the fact that it allows learners to play an active role in their own measurement/evaluation and in the process to facilitate opportunities for student empowerment. It also helps to counter some of the serious disadvantages of conventional examinations, for example the effect of emotional disturbances on student performance, the great stressful conditions, and so forth (see chapter 3, section 3.5.1.2). Personal and performance records also help to create a more complete picture of learners, their backgrounds and former achievements. With regards to apprenticeships/internships no further elaboration is necessary in addition to what was mentioned in section 4.5.5.3. It can be emphasised again that the implementation of apprenticeship/internship programmes at TSA in BM (at least on third year level) will be a very sensible step to take.

4.6 A BRIEF SCRUTINY OF COMPUTER-BASED MEASUREMENT

Because of the strong emphasis placed on technology-enhanced learning and the fact that TSA started developing the infrastructure for the establishment of a virtual campus, the researcher has decided to examine computer-based evaluation as one of the alternatives that might also be considered for evaluation purposes in Business Management.

Computer-based evaluation is singled out from the scope of technological instruments, because it is apparently used at an increasing rate at distance education institutions in particular (Crock, Dekkers & Cuskelly, 1994: 133). It is also very clear that TSA’s virtual campus (see section 2.4.3, chapter 2) will use the computer for all those
purposes indicated. It should however be noted that the most common test or examination type used in computer-based measurement is the objective-type question (Merrill et al, 1986: 220; Alessi & Trollip, 1991: 204). Since the objective-type question was already examined in section 3.3.2.1, chapter 3, this section will only focus on an examination of the computer as a measurement instrument.

4.6.1 Description of computer-based measurement

Alessi & Trollip (1991: 205) state that there are two main ways in which computers can be used in the measurement/evaluation process:

- First, by using the computer for test construction;

- Second, by using the computer to administer tests.

Merrill et al (1986: 216 - 220) add another few uses of the computer in the measurement process, that is:

- First, computers can also be used for the scoring of tests by means of special answer sheets marked by optical mark readers.

- Second, computers can also be used by learners to do their tests on the computers, that is the so-called terminal-based computer measurement.

It should be noted that, although reference is only made to tests, which is a formative measurement instrument, computers can also be used for summative measurement purposes.

With this brief description of the ways in which computers can be used in the measurement process, the advantages and disadvantages of computer-based measurement will be considered next.
4.6.2 Advantages of computer-based measurement

In terms of the generating of item-banks or question-pools on a large scale, that is instances where groups of educators work together to generate new test items, Alessi & Trollip (1991: 207) identified the following advantages of computer-based measurement:

- The fact that a number of educators help to develop new questions, creates the opportunity for individual educators to have access to a larger number of questions than would normally have been the case.

- The greater selection of questions allows educators a much greater source for new test questions without having to fear that learners will learn to know all the possible questions available.

- The fact that various educators would develop items for the pool, will lighten the burden on individual educators to produce new questions.

- The opportunity for regular evaluation of questions would also be created.

Alessi & Trollip (1991: 208) also identified the following general advantages of computer-based measurement:

- The measurement process can become very individualised which, in turn, will allow learners to do tests at their own pace.

- Tests can be adapted to suit the specific approaches or needs of individual learners.

- Educators will have the opportunity to store the responses of learners for the purpose of evaluation or improving the questions.
• The storing of learner responses can also enable educators to obtain valuable information about their learners.

Merrill et al (1986: 216 - 221) also mention the same advantages as Alessi & Trollip. However, they do mention a few additional advantages. Those advantages are the following:

• Terminal-based measurement ensures a much simpler logistical process than normal measurement, because there is little need for paperwork such as examination papers, answer sheets and so forth.

• Learners become more active participants in the teaching/learning process and this promotes greater independence from educators.

A few additional advantages that the researcher can foresee, in particular with regard to distance education, is the fact that computer-based measurement can speed up the process of courseware delivery to learners and it creates the opportunity for almost immediate feedback to learners on their progress and achievements.

4.6.3 Disadvantages of computer-based measurement

Merrill et al (1986: 218 - 221) and Alessi & Trollip (1991: 207 - 208) identified the following disadvantages of computer-based measurement:

• The generation of test items for a small number of learners is not a very cost-effective and economic process. More items than necessary for one test have to be generated and this can be a very time-consuming process.

• Individualised measurement programmes can cause serious logistical problems, for example:
  - It becomes difficult for educators to keep track of the progress of all learners if they are progressing at different rates.
- It can become a difficult exercise to maintain test security in cases where learners are allowed to repeat tests more than once.

- In the case of terminal-based measurement, a need will probably arise for a large number of computers to be made available to learners. Such a need can have serious cost implications for an institution.

- In cases where an institution does not have enough computers available, it can cause long waiting periods for different groups of learners.

- It is very difficult to score essay-type questions via computer. Computer-based measurement seems to be more suitable for objective-type questions.

- The items that are easily scorable by computer, are not necessarily the best to use for the evaluation of the mastery of particular skills.

- It is also possible that the system can be abused by some users, for example:
  - Educators can sometimes use questions merely for the sake of convenience and not be concerned about the effectiveness of those questions.
  - Questions are also sometimes used before they were tested for validity and suitability.

This list of disadvantages is not exhaustive, but it does include some of the main disadvantages that can be associated with computer-based measurement. It would however seem as if the major advantages of computer-based measurement overshadow its disadvantages.

4.6.4 Suitability for TSA conditions

Computer-based or on-line instruction and measurement, will have some major
advantages for a distance education institution such as TSA, as pointed out in section 2.4.3, chapter 2. Below is a summary of some of the main advantages for TSA:

- Computer-based instruction will speed up the process of courseware delivery.
- It will create the opportunity for individualised instruction and evaluation.
- It will provide learners with almost immediate feedback.
- For TSA with her enormous learner population, it will make the process of formative measurement much easier.

However, the major disadvantage of the computer-base measurement, as in the case of objective-type questions, is that it cannot easily provide evaluation on the higher cognitive levels. In particular, the aspect of application skills does not seem possible via computer-based measurement.

4.6.5 Suitability for BM

On the basis of the types of questions used in BM, see section 5.3, chapter 5) it would appear that computer-based measurement will be more suitable for BM I because the multiple-choice question is currently used only in BM I. Only in the case where lecturers in BM II and III also decide to use the multiple-choice question, would computer-based measurement in its current format become useful to them. Such a move by BM II and III lecturers would however be in contradiction to the principles and prescriptions of SERTEC, see section 5.5, chapter 5, because as stated, it is very difficult to use multiple-choice questions for the fostering of application skills.

4.6.6 Financial implications

The implementation of computer-based measurement can be a costly process, in particular where large numbers of computers have to be acquired to accommodate
large numbers of learners who do not have access to computers. TSA, who embarked on a decentralisation exercise a few years ago, has various regional offices and study centres in all the provinces in South Africa. To provide all those study centres with enough computers will be costly. What might help is the fact that TSA also joined the Cyber Connection who offered free access to COLISA learners on the internet (Van der Merwe & Lazenby, 1997:1). Coupled with this is also TSA’s partnership in the COLISA agreement that might help to reduce individual expenses because of the possibility of shared costs. COLISA stands for: Confederation of Open Learning Institutions in Southern Africa, and consists of TSA, UNISA and Vista University.

With these few comments, an overall conclusion of this chapter will be presented at this stage.

4.7 CONCLUSION

After a careful scrutiny of the full contents of this chapter, the researcher reached the following conclusions:

First is the fact that actual performance measurement, despite a number of disadvantages or obstacles, appears to be the most ideal form of measurement for BM at TSA. However, to realise such an aim will require from TSA to introduce apprenticeship/internship programmes (at least on third year level).

The second conclusion is that simulations, and in particular management exercises, also represent a very useful and effective form of performance-based measurement for BM at TSA which should be implemented without delay.

The third conclusion reached in this chapter is that many of the alternative instruments discussed, are very continuous or formative in nature (for example supervisor ratings, peer ratings, self-evaluation, portfolios, profiles, records of achievement, and so forth). A possible implication of this situation is that formative evaluation in essence might also be a possible alternative to conventional summative evaluation instruments. A
total move away from summative evaluation is not something new because in the US school system the final examination has already been abolished for the sake of continuous/formative evaluation (see section 3.4.1.1).

A move towards formative evaluation can have many advantages, for example:

- It will help to eliminate the problem of examination stress.
- It will eliminate the criticism that learners are required to give account of a full year's work in a three hour exam paper.
- It will give a more realistic manifestation of students' everyday performance and skills.

The final conclusion reached in this chapter is that time, cost and standardisation factors appear to be the biggest obstacles in the introduction and implementation of performance-based measurement instruments. However, the researcher's view on this matter is that, in the light of the numerous advantages and benefits of performance-based measurement, a full-scale research project should be launched to find ways in which these obstacles can be overcome or neutralised in order to effect the use of performance-based measurement.

The main question that will have to be answered in the next chapter, is whether these obstacles (costs and time) should be allowed to prevent the implementation of performance-based measurement instruments in BM at TSA. In order to find the answer to that question the next chapter will be devoted to an actual consultation with all stakeholders involved in BM at TSA. This will include learners in BM I, II and III; lecturers, departmental heads, senior management, industry and all other affected parties.

With these conclusions as a guideline to the rest of this thesis, a brief summary of this entire chapter will be provided.
4.8 SUMMARY

In this chapter an attempt was made to supply a description and explanation of the meaning and scope of alternative educational measurement instruments. A substantiation was given why the term, performance-based measurement, would be used to represent the domain of alternative measurement instruments. Thereafter particular aspects regarding performance-based measurement were discussed. This included the following: A definition of performance-based measurement; the various types of performance-measurement; the characteristics, validity, reliability and accountability of performance-based measurement and its general advantages and disadvantages. Thereafter a brief evaluation of different performance-measurement categories was undertaken. The categories included: actual performance measurement, simulations, observational measurement, oral measurement and programme requirements.

The next chapter will involve empirical research in which all relevant stakeholders in BM at TSA will be interviewed and their responses be analysed in order to reach a final conclusion regarding the implementation of performance-based measurement instruments for BM at TSA.
CHAPTER 5

BUSINESS MANAGEMENT CURRICULUM VERSUS MEASUREMENT INSTRUMENTS

5.1 INTRODUCTION

The previous chapter dealt with an examination of available alternative measurement instruments. In that chapter (chapter 4) it was established that most of the different concepts referring to alternative evaluation instruments, have one main similarity in common, namely the aspect of performance. As a result, the concept of performance-based measurement instruments was examined in detail. Aspects that were examined are: A definition of performance-based measurement, its advantages and disadvantages; the aspects of validity, reliability and accountability. Following this, a wide variety of performance-based measurement instruments were examined critically in terms of its advantages, disadvantages and suitability for TSA conditions. In conclusion, it was stated that actual performance measurement and simulations appear to be the most suitable measurement instruments that can serve as alternatives to the conventional standard written examination.

In this chapter the aim is to analyse the curriculum of BM and the measurement instruments used in BM to determine whether there is indeed a positive correlation between curriculum objectives and the measurement instruments. The main question that needs to be answered, is: Do the measurement instruments in BM really measure those aspects intended to be measured by the curriculum objectives?

5.2 RESEARCH METHODOLOGY

Some of the information in this chapter was obtained by means of a survey and document review or content analysis. Although it seems as if the survey is a method commonly used in quantitative research (Lancy, 1993:12), Marshall & Rossman (1995:86) describe it as one of the supplementary data collection techniques that can
be used in qualitative research. Document review on the other hand, is one of the primary data collection techniques that can be used in qualitative research (Marshall & Rossman, 1995: 78). However, since the survey and document review are not the main instruments used, they will be described briefly.

5.2.1 Survey and questionnaire

Although the survey is regarded as a supplementary data collection instrument, it is discussed first because it was the first method used by the researcher. A general definition of the verb “survey” is as follows:

- “take general view of, form general idea of the arrangement and chief features of” (Sykes, 1982: 1075).

This definition accurately expresses the purpose of the survey in this research.

Marshall & Rossman (1995:96) describe the overall aim of research by means of surveys to provide statistical descriptions or explanations of the “variability” of particular characteristics of a population. They also state that the survey is a recommended research instrument in instances where a small amount of data is required from a large population.

Surveys can be done either by mail, telephone or personal interview and are normally conducted by means of questionnaires. In this research, a survey was done by mailing questionnaires to the relevant research population.

Marshall & Rossman (1995: 95, 96) describe the purpose of questionnaires as that of gathering information about certain characteristics, attitudes or beliefs of a sample of a particular population. Questionnaires are normally used to obtain data on aspects that can be measured fairly accurately by means of self-report. Marshall & Rossman point out that questionnaires place researchers in a difficult position, because they must
assume that responses received, reflect an honest and accurate report by participants (1995 : 96). However, this does not totally destroy the usefulness and value of questionnaires. Questionnaires can be structured, but also contain close-ended and open-ended questions. The questions should however be scrutinised carefully for bias, sequence, clarity and face validity (Marshall & Rossman, 1995 : 96).

5.2.2 Document review

Marshall & Rossman (1995 :85) describe the review of documents as an unobtrusive data collection instrument. Such document reviews can help to establish a relatively good understanding about the values and beliefs of participants. A variety of documents in fact can be taken for review. Marshall & Rossman provide the following examples:

- Minutes of meetings
- Logs
- Announcements
- Formal policy statements
- Letters
- Marriage records
- Budgetary allocations, and so forth (1995 : 85)

Marshall & Rossman also distinguish a particular type of document review, known as "content analysis". Content analysis is concerned with the systematic scrutiny of all types of documents and the objective documentation of the information obtained. Although content analysis is concerned mainly with written materials, it can also be used for other types of communication, for example music, pictures, speeches, and so forth (Marshall & Rossman, 1995 : 86).

One of the main advantages of content analysis is that it is an unobtrusive research instrument and causes no disturbance to settings. A researcher will be able to decide on
the basis of the data what aspects deserve the main emphasis. Content analysis can also be evaluated by an outsider regarding the validity of facts and the care taken by the researcher.

In the case of this research, the researcher applied the method of content analysis on the curriculum of BM and a number of examination papers in BM I, II and II. The main aims with this analysis was to determine whether the measurement instruments used in BM for the evaluation of learner progress, did in fact correlate with the curriculum objectives of BM. In other words, it was to determine whether the measurement instruments used on each year level would indeed help realise the curriculum objectives.

5.3 MEASUREMENT INSTRUMENTS IN BM AT SA TERTIARY EDUCATION INSTITUTIONS

A short questionnaire was sent to the Business Management Departments of a number of tertiary institutions in South Africa, which involved 20 universities and 14 technikons. A copy of this questionnaire is supplied in annexure A. 13 questionnaires were received from the universities and 11 from the technikons. Unfortunately the responses returned by University of the Witwatersrand and Potchefstroom University for Christian Higher Education could not be used because they had not been completed by the BM subject lecturers or departments of those institutions. This meant 11 were received from universities or 55% and 11 from the technikons or 78,5%. From a total of 34 institutions approached this gives a total of 22, which is 64,7%. Although the information obtained from these questionnaires will not be used to generalize, a number of conclusions can be drawn. Before discussing a summary of such observations, the main forms of measurement instruments at these institutions, both summative and formative, will be mentioned. The institutions are supplied in random order. (Please note, the respondents to the questionnaires gave their permission for reporting this information in this way.)
Diagram 1: Summative measurement instruments at some universities in South Africa

<table>
<thead>
<tr>
<th>University</th>
<th>Summative measurement instrument</th>
<th>Nature of instrument</th>
<th>Learner year level</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of the Western Cape</td>
<td>Written examinations</td>
<td>Short questions</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td>Oral examinations (for borderline cases after written examination)</td>
<td>Multiple-choice Essay-type</td>
<td>I, II, III</td>
</tr>
<tr>
<td>University of Natal (Pietermaritzburg)</td>
<td>Written examinations</td>
<td>Short questions</td>
<td>II, III</td>
</tr>
<tr>
<td></td>
<td>Open-book Written case studies</td>
<td>Multiple-choice Essay-type</td>
<td>II, III</td>
</tr>
<tr>
<td></td>
<td>Written and practical examinations</td>
<td></td>
<td>III</td>
</tr>
<tr>
<td>University of Zululand</td>
<td>Written examinations</td>
<td>Short questions</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple-choice Essay-type</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>II, III</td>
</tr>
<tr>
<td>University of Pretoria</td>
<td>Written examinations</td>
<td>Structured essay-type</td>
<td>I</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>----</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Written case studies</td>
<td>III</td>
</tr>
<tr>
<td>Oral examinations</td>
<td></td>
<td>Short questions</td>
<td>I, II, III</td>
</tr>
<tr>
<td>(in special circumstances)</td>
<td></td>
<td>Multiple-choice</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Essay-type</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Structured essay-type</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Written case studies</td>
<td>II, III</td>
</tr>
<tr>
<td>Computer-based</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>examinations</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>University of Port Elizabeth</th>
<th>Written examinations</th>
<th>Short questions</th>
<th>I, II, III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Essay-type</td>
<td>III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Written case studies</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Oral examinations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(only in special cases)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computer-based</td>
<td></td>
<td>Multiple-choice</td>
<td>I, II, III</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>University of the Orange Free State</th>
<th>Written examinations</th>
<th>Short questions</th>
<th>I, II, III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Written examinations</td>
<td>Multiple-choice</td>
<td>I, II</td>
<td></td>
</tr>
<tr>
<td>Written and practical examinations</td>
<td>Essay-type</td>
<td>I, II, III</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Structured essay-type</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Written case studies</td>
<td></td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rand Afrikaans University</td>
<td>Written examinations</td>
<td>Short questions</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple-choice</td>
<td>I, II</td>
</tr>
<tr>
<td></td>
<td>Essay-type</td>
<td>I, II, III</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University of South Africa</td>
<td>Written examinations</td>
<td>Short questions</td>
<td>II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple-choice</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td>Essay-type</td>
<td>II, III</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Written case studies</td>
<td></td>
<td>III</td>
</tr>
<tr>
<td>Oral examinations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(only in special cases)</td>
<td></td>
<td></td>
<td>I, II, III</td>
</tr>
<tr>
<td>Computer-based</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technikon</td>
<td>Measurement instrument</td>
<td>Format of measurement instrument</td>
<td>Learner year level</td>
</tr>
<tr>
<td>---------------------------</td>
<td>------------------------</td>
<td>-----------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Technikon Pretoria</td>
<td>Written examination</td>
<td>Short questions</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple-choice</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Essay-type</td>
<td>II</td>
</tr>
<tr>
<td>Institution</td>
<td>Written examination</td>
<td>Structured essay-type</td>
<td>Essay-type question</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------</td>
<td>-----------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Technikon Orange Free State</td>
<td>Written examination</td>
<td>Short question</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Essay-type question</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Written case studies</td>
<td>III</td>
</tr>
<tr>
<td>Port Elizabeth Technikon</td>
<td>Written examination</td>
<td>Short questions</td>
<td>I, II</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple-choice</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Essay-type question</td>
<td>II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Written case studies</td>
<td>III</td>
</tr>
<tr>
<td>Mangosuthu Technikon</td>
<td>Written examination</td>
<td>Short questions</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Essay-type question</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Written case studies</td>
<td>II, III</td>
</tr>
<tr>
<td>Technikon Natal</td>
<td>Written examination</td>
<td>Short questions</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Essay-type question</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Written case studies</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Setlogelo Technikon</td>
<td>Written examination</td>
<td>Short questions</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple-choice</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Essay-type question</td>
<td>I, II, III</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Written case studies</td>
<td>III</td>
</tr>
<tr>
<td></td>
<td>Oral examination</td>
<td></td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>Written and</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Institution</td>
<td>Type of Exam</td>
<td>Question Types</td>
<td>Sections</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>--------------------------------------------------</td>
<td>-------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>M.L. Sultan Technikon</td>
<td>Written examination</td>
<td>Short questions, Multiple-choice, Essay-type question, Written case studies</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Transkei Technikon</td>
<td>Written examination</td>
<td>Short questions, Multiple-choice, Essay-type question, Structured essay-type</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Northern Transvaal Technikon</td>
<td>Written examination</td>
<td>Short questions, Multiple-choice, Essay-type question, Structured essay-type, Written case studies</td>
<td>I, II, III</td>
</tr>
<tr>
<td>Peninsula Technikon</td>
<td>Written examination</td>
<td>Short questions, Multiple-choice, Essay-type question, Structured essay-type, Written case studies</td>
<td>I, II, III</td>
</tr>
</tbody>
</table>

The table lists various types of examinations and the corresponding question types and sections for different institutions.
Based on these diagrams, a number of observations can be made. A summary of such observations follows:

- It would appear as if written examinations are still the dominant form of measurement in many tertiary educational institutions in South Africa.

- Short questions, multiple-choice questions, essay-type questions, structured essay-type questions and written case studies are apparently in regular use at a number of South African tertiary educational institutions.

- The written case study format is a common format of BM examination papers in many tertiary educational institutions in South Africa.

- A number of tertiary educational institutions in South Africa use computer-based forms of measurement in BM which can possibly signify a move towards alternative forms of measurement.

- The open-book type of examination is not yet in regular use in South African tertiary education at undergraduate level for BM (only at one university and one technikon in this study).
• Oral examinations are still used at some tertiary institutions in South Africa, but only for special cases, e.g. borderline cases, and only as a follow-up on the written examination.

• A few tertiary educational institutions in South Africa use both written and practical, unspecified, measurement instruments for evaluation purposes in BM.

In conclusion, it would appear that the most dominant form of measurement in BM at tertiary educational institutions in South Africa is still the written examination. However, there seems to be a move towards alternative forms of evaluation other than the written examination, e.g. computer-based evaluation, at some tertiary educational institutions in South Africa. It also seems as if South African tertiary education still has a long way to go in terms of the common use of alternative forms of measurement.

The researcher will at this stage do a brief scrutiny of the specific measurement instruments used for evaluation purposes in BM at Technikon SA.

5.4 MEASUREMENT INSTRUMENTS IN BM AT TSA

Based on the information provided in diagram 2, it seems as if BM at undergraduate level at TSA complies with the ordinary or conventional forms of measurement used at other tertiary educational institutions in South Africa. The following observations on the unique situation in BM at TSA can be made:

• The multiple-choice question is only used at first-year level.
• At all three levels (I, II, III) short type of questions are in common use.
• The written case study is also used on all three levels.
• Although the structured essay-type question is used at all three levels, the general essay-type question is used only on third year level.
• The open-book format is used only in BM III and this makes TSA the only technikon on the list supplied in diagram 2 to use the open-book examination in BM.
• However, although a move towards alternative measurement instruments is being made, as can be noticed in the use of the open-book format, the written examination is still the only summative form of measurement in BM at undergraduate level at Technikon SA.

The last observation is one of the main factors that contributed to this research, because it was evident, as will be seen in chapter 4, that there are a variety of alternative forms of measurement which create fairer conditions for learners in any subject. It has already been pointed out that a number of educationalists (Brown & Knight, 1994:44; Justiz & Kameen, 1989:40; Heywood, 1989:19, 86) recommend the use of a variety of measurement instruments in the evaluation process, particularly so in a multicultural and multilingual context such as South Africa.

This brief description of the use of summative measurement instruments in South African tertiary education and at TSA, necessitates a brief evaluation of the main summative measurement instruments described in the first half of this chapter.

5.5 SERTEC PRINCIPLES CONCERNING THE CURRICULUM

The SERTEC manual (1996: 6 - 8) lists a number of characteristics of technikon instruction in general. These characteristics were obtained from the NATED 02-118 (88/07) report. These characteristics can however be regarded as some of the objectives technikon education programmes should strive towards. Some of the most significant characteristics are the following:

• Technikon education should also be formative in nature.
• There is normally a great emphasis on the fostering of a proper understanding of technology and of occupations. The emphasis is more on such understanding, than on academic and scientific knowledge.
• Technikon education should be aimed towards immediate application of skills. In other words, after receiving their technikon qualifications, learners should be "ready" and fully oriented in terms of the skills required from them.
When designing and developing new education programmes, technikons must consult industry and address the requirements of industry in that programme. TSA in particular lays strong emphasis on this industry liaison.

The focus of technikon instruction programmes is more on careers than on disciplines (SERTEC, 1996: 6, 7).

The SERTEC manual provides clear guidelines on the nature and contents of technikon instructional programmes. The focus is clearly on developing the skills of application of knowledge in practical situations. The following prescriptions regarding the nature of contents are stated:

- “A maximum of 20% of the credits for the so-called A-type subject contents.
- A minimum of 60% and a maximum of 80% of the credits for the so-called B-type subject content.
- A maximum of 20% of the credits for the so-called C-type subject content.” (SERTEC, 1996: 35).

A-type subject contents consist of the basic psycho-motor skills that learners should have mastered in order to perform higher level skills.

B-type subject contents are those contents that focus specifically on the application of subject knowledge or technology in new or practical situations.

C-type subject contents are more theoretically and scientifically oriented.

It becomes clear that the emphasis in technikon education programmes is primarily an application skills.

The SERTEC manual also encourages the introduction of experiential learning at technikons (SERTEC, 1996: 35). Experiential learning entails the placement of learners in a certain workplace for periods of time in order to gain the actual practical skills that such an industry normally requires from employees.
Because it can sometimes be difficult to arrange such opportunities for experimental instruction for a large number of learners, SERTEC limits the weight of the mark for experiential learning to 33.33% of the credits for a technikon qualification (SERTEC, 1996: 36).

Based on this information, it is clear that the practical application and actual skills that will be required in the workplace, should be a primary aim of all technikon programmes.

5.6 THE BM CURRICULUM VERSUS MEASUREMENT INSTRUMENTS

An attempt was made to analyse the aims and objectives of the BM curriculum and to compare those objectives with the evaluation methods used to test the achievement of those aims and objectives. However, the focus in the analysis was not so much on subject contents, but on the curriculum objectives and measurement instruments. The following process was followed:

• First, the curricula of BM I, II, and III were analysed to determine their exact objectives and what those objectives entail.

• Second, the measurement instruments used in BM I, II and III were analysed to determine exactly what they intended to measure.

• Third, a comparison was drawn between the curriculum objectives of BM I, II and III, and the measurement instruments used to determine whether a positive correlation between curriculum objectives and measurement instruments exists in all three cases.

Below is a description of the findings of this process.
5.6.1 Findings from the BM I curriculum and measurement instruments

5.6.1.1 Findings from the analysis of the BM I curriculum

In BM I no clear objectives seem to have been stated. Only the subject contents to be taught to learners were described and the length of the examination paper to be written.

From an analysis of the BM I curriculum it seems as if the focus is primarily on the introduction of the subject contents and fundamental subject terms to the first year learner. There is a great emphasis on the definition of concepts and fostering an understanding of these definitions.

Examples of introductory statements:

- Introduction to management in general (p. 2).
- Introduction to planning (p. 2).
- Introduction to organizing (p. 2).
- Introduction to leading (p. 2).
- Introduction to control (p. 3).
- Introduction (The Human Resources Function) (p. 4).
- Introduction to the purchasing function (p. 5).

Examples of definitions:

- Definition and description (of the concept of BM) (p. 1).
- Definition of Enterprise (p. 1).
- Definition of management (p. 2).
- Definition and objectives (of the Administrative Function) (p. 3).
• Definition (of the Production Function) (p. 6).

• Definition of external relations (p. 7).

• Definition of the concept (Market segmentation) (p. 7).

• Definition of a small business (p. 8).

The only instance where direct reference was made to what learners should be able to do (outcomes), was the following:

• “Learners must understand that administrative management must take place in each business function” : (p. 4) (BM I Curriculum, 1997: 1 - 8).

On the basis of the above information it would seem that some of the main aims and objectives of the BM I curriculum are the following:

• To introduce the first-year BM learner to the nature of the subject.

• To enable learners to know and be able to define certain basic subject concepts.

• To establish a basic understanding among BM I learners of essential subject contents.

5.6.1.2 Findings from an analysis of BM I measurement instruments

A scrutiny was conducted of the types of questions in the examinations learners have to write at the end of the year. The following four BM I examination papers were taken for analysis:

• November 1995

• January 1996

• May 1996

• November 1996
As indicated in diagram 2 the following types of questions apparently occur regularly in BMI examination papers:

- The multiple-choice questions
- Short questions (questions counting approximately 2 - 5 marks)
- The structured essay-type question (to a lesser extent)
- The case study approach

A further analysis was done to determine what main action (key) verbs were normally used in these questions. Action verbs help to give an understanding of the cognitive levels, according to Bloom's taxonomy, of questions. The researcher is fully aware that a particular action verb is not bound to a particular cognitive level, and that it is in fact the context in which a particular verb is used, that will determine its cognitive level. However, the cognitive level indicated next to each verb in this chapter, indicates the cognitive level that these verbs normally represent. The researcher compared these cognitive levels to the contexts in which these verbs were used in the examination papers and there seems to be a positive correlation.

Some general action verbs used in the examination papers are the following (Next to each verb the estimated cognitive level of that verb is provided):

- The multiple-choice questions were almost without exception on the knowledge level.

- The other questions used the following action verbs:
  
  - explain: comprehension
  - compare: comprehension
  - distinguish: comprehension
  - define: knowledge
  - be practical: application
  - provide reasons, motivate: comprehension
• indicate  
  comprehension
• discuss  
  comprehension
• what type  
  knowledge
• why  
  knowledge
• name / list  
  knowledge
• examples  
  application
• identify  
  knowledge
• give your opinion  
  evaluation

The action verbs most frequently used seem to be:

• explain
• discuss
• list.

These action verbs appeared most commonly in all three papers. Interestingly, the verb "define" is also used, but on a less frequent basis. On occasion these action verbs were followed by:

• “Motivate your answer”
• “Give reasons”
• “Give practical examples”

The cognitive levels on which each of the dominant verbs are normally used are as follows:

• explain = comprehension
• discuss = comprehension
• list = knowledge
5.6.1.3 Findings from a comparison between BM I curriculum objectives and measurement instruments

On the basis of the above analysis it seems as if the main objectives of the questions used in the measurement instruments are to test or determine the level of knowledge and understanding learners have gained from the subject contents. This conclusion correlates with the assumed objectives of the curriculum as stated in section 5.6.1.1.

In conclusion then, it can be stated that the measurement instruments used in BM I are indeed aimed at determining to what extent the objectives of the curriculum were achieved. The types of questions used can also be regarded as correlating with the objectives of the BM I curriculum.

5.6.1.4 Findings from a comparison of BM I curriculum objectives and measurement instruments with SERTEC prescriptions

A serious question that did however arise, is whether the curriculum objectives and measurement instruments of the BM I curriculum correlate with SERTEC requirements and principles for courses offered on higher education level (see section 5.5)? In the SERTEC prescriptions it is clear that the greatest emphasis is placed on the development of application skills among learners.

It does seem as if attempts were made in the BM I examination papers to promote or facilitate application. Most of the questions are based on written case studies. However, in most cases it appears as if the case studies merely serve to act as a starting point, because the answers required from learners would in many cases entail a reproduction of the knowledge and contents of the subject.

The other problem also regarding the use of written case studies, is that it is still a written form of measurement. The question arises, whether it is possible to use written examinations to develop and measure skills of a practical nature? The written case study seem to suffers from the same shortcomings as the conventional written
examination in as far as the aspect of application is concerned (see section 3.5.1.2, chapter 3).

It seems that, although attempts were made to promote greater application of knowledge and skills among learners, the methods used for that purpose suffer from certain shortcomings. Also, no curriculum objectives particularly aims at the development of application skills in learners.

In conclusion, it seems that there are certain shortcomings in the BM I curriculum and measurement instruments with regard to the SERTEC prescriptions.

5.6.2 Findings from the BM II curriculum and measurement instruments

5.6.2.1 Findings from the analysis of the BM II curriculum

The BM II curriculum also involves the introduction of new concepts to learners, possibly those that were not touched on during the first year, and the definition of a number of new concepts. However, in this curriculum clear objectives specifying the desired outcomes, were stated.

It appears as if the BM II curriculum consists of the following three main functions:

- The production function
- The public relations function
- The financial function

- Objectives of the production function

In the case of the production function no clearly formulated aims or objectives are stated. However, it seems as if similar objectives as in the case of BM I were also intended in this instance. There is for example also a focus on the definition of new concepts, for example:
• Definition of the production function

• Definition of work study (BM II Curriculum, 1997: 11, 12).

The assumption reached by the researcher is that the main objectives of this function is to establish an adequate level of knowledge and understanding of the function among learners.

• Objectives of the public relations function:

The exact formulation of the objectives of this function is as follows:

"After studying the public relations function the learner should:

• Clearly understand the role of public relations in the firm;
• understand the communication process and have knowledge of factors that influence communication;
• know the public relations process and therewith the day-to-day activities of the public relations department;
• be aware of the tools used in public relations and their strengths and limitations, the skills required to handle these tools and the practical implications of using them;
• be aware of the different public relations media, the appropriateness of the various media to particular audiences and communications needs, what the media demand from the firm and the cost involved;
• understand the interface between public relations and the other functions, appreciate the practical value of public relations to his own and other functions and to realise what inputs public relations in turn require to make an optimum contribution to the success of the business" (BM II Curriculum, 1997:12, 13).

• Objectives of the financial function:

These objectives are:
• "To provide insight into those principles of financial management that are considered important to financial practice”
• "To be able to apply these principles”.

On analysing the key terms (action verbs) used in each of these objectives, the following was found:

• In terms of the production function the emphasis is on knowledge and understanding.
• In terms of the public relations function the emphasis is on knowledge, understanding, awareness, appreciation and realisation of certain factors or contents.
• In terms of the financial function the emphasis is on understanding and application.

It is significant to note that it is only with regard to the financial function where the establishment of application skills is clearly stated as an objective of the curriculum.

5.6.2.2 Findings from an analysis of BM II measurement instruments

The following examination papers for BM II were taken for analysis:

- November 1994
- November 1995
- May 1996
- November 1996

The main question types are those indicated in diagram 2. These are:

• Short questions
• Structured essay-type questions
The written case study approach was also followed in all four examination papers.

The key action verbs most commonly used were the following:

- What type
- What is
- Motivate/give reasons
- Discuss
- Define
- Refer to
- Explain
- Own examples
- Identify
- Be practical (give practical examples)
- Why
- Indicate
- List
- Recommend
- Would you say....
- Design
- Take into account
- Decide on

An analysis of the BM II examination papers on the basis of key (action) verbs shows the following:

The most commonly used verbs are:

- explain
- discuss
- list.
Verbs that were occasionally used in conjunction with the above three are:

- motivate
- provide reasons
- provide examples
- be practical.

It appears as if the purpose with these verbs used in conjunction with the main ones, was to promote application, or a practical approach among learners.

The main verbs, however, appears to correlate with the ones used in BM I and also fall on the cognitive levels of knowledge and understanding.

It would however be unfair not to mention that attempts were apparently made to use action verbs at a higher cognitive level, especially in the May and November '96 papers. However, it happened in isolated cases only. Examples of such higher cognitive verbs are:

- Recommend = evaluation
- Would you say... = evaluation
- Design = synthesis

It seems therefore as if the objectives of the main questions were, as in the case of first year level, to determine the level of understanding and knowledge learners have attained with regard to subject contents and skills. In addition to this, it appears as if a much greater emphasis was placed upon the application facet than in the case of BM I. However it was again a case of application within the limitations of the written word.
5.6.2.3 Findings from a comparison between BM II curriculum objectives and measurement instruments

In a comparison between the curriculum objectives and the objectives of the examination questions, it seems as if there is also a positive correlation between those objectives, for example:

- Both sets of objectives aim towards the fostering of adequate knowledge and understanding among learners.
- The element of application skills is included in both.

In conclusion, it would therefore seem as if the measurement instruments used in BM II are aiming at determining to what extent the curriculum objectives of BM II were realised.

5.6.2.4 Findings from a comparison of BM II curriculum objectives and measurement instruments with SERTEC prescriptions

The same findings in section 5.6.1.4 were also reached after a comparison of the BM II curriculum and measurement instruments with the SERTEC prescriptions and the same comments will therefore apply.

5.6.3 Findings from the BM III curriculum and measurement instruments

5.6.3.1 Findings from an analysis of the BM III curriculum

The main content areas included in the BM III curriculum are the following:

- The Human Resources Function
- The Marketing Function
- The Purchasing Function
- The Management Function (BM III Curriculum, 1997: 20)
The curriculum objectives of each of these functions are clearly stated in the curriculum. These objectives are as follows:

- Objectives of the human resources function:

No overall goal is stated here. The main objectives are the following:

"After studying the personnel function, the learner should be able to:

- apply the management functions, namely planning, organizing, leading and control, to the manpower in his department;
- perform and interpret a job analysis so that he knows what his subordinates are supposed to do;
- perform his own recruitment, selection and placement of personnel;
- develop a comprehensive induction programme and to implement it with all newcomers;
- manage the training and development of his staff himself;
- distinguish between a salary notch and a salary scale;
- complete the performance appraisals of his subordinates and use these for their career development;
- monitor contract administration;
- identify unsafe and unhygienic conditions;
- follow the procedures for promotions, transfers, resignations, retirements and dismissals." (BM III Curriculum, 1997: 20, 21).

The action verbs used in the curriculum objectives are:

- apply
- perform
- interpret
- develop
- implement
• manage
• distinguish between
• complete
• monitor
• identify
• follow procedures

• Objectives of the marketing function

The overall goal of the marketing function places the emphasis on understanding.

The stated objectives for this function are:

"After completing this syllabus, learners should:

• understand the nature of marketing as distinct from, for example selling;
• gain an insight into factors that influence consumers in their behaviour towards the firm and its offerings;
• have developed an instinctive appreciation for the centrality of meeting customers' needs, both to the marketing function and to the firm as a whole, through all of its functions; and
• appreciate the interrelationships between the functions of the business, and especially that of the learner's specialist derivation.” (BM III Curriculum, 1997: 23).

The main action verbs used in these objectives are the following:

• understand / insight
• appreciate

• Objectives of the purchasing function
The goal places the emphasis on understanding and implementation. The objectives are the following:

"After studying the purchasing function, learners should be able to:

- give a brief overview of the historical development of the purchasing function;
- define the purchasing function and formulate objectives;
- give an exposition of the importance of the purchasing function to the enterprise;
- understand and explain managerial activities as applicable to the purchasing function;
- explain the purchasing cycle; and
- explain and apply the elements of the purchasing procedure." (BM III Curriculum, 1997: 26)

The action verbs used in the objectives are:

- give an overview
- define
- give an exposition of
- understand
- explain
- apply

• Objectives of the management function

The goal places the emphasis on understanding and application. The objectives of this function are:

"After studying the management function, learners should be able to:

- distinguish between strategic, functional and operations management; and
• discuss the nature and essence of management, strategic management, functional management and operations management.” (BM III Curriculum, 1997: 29).

The main action verbs used in the objectives are:

• distinguish
• discuss

After scrutinising these goals and objectives, it became clear to the researcher that the emphasis in the objectives of the BM III curriculum ranges from the knowledge level (define) to the evaluation level (implement, develop). A strong emphasis is placed on application and this correlates strongly with the prescriptions of SERTEC, section 5.5.

The findings of the analysis of BM III measurement instruments will be discussed next.

5.6.3.2 Findings from an analysis of BM III measurement instruments

The BM III curriculum gives the following prescriptions regarding the measurement instruments to be used:

“Examination:
  a) Comprises of one third of a three-hour paper,
  b) The second paper may take the form of a case study or any other assignment involving practical work. This practical component may incorporate the work of all three years” (BM III Curriculum, 1997:20).

These prescriptions are made with regard to all the functions mentioned in 5.6.3.1 (The Human Resources Function, The Marketing Function, The Purchasing Function, The Management Function).

The measurement instruments used in BM III are the following:
• Short questions
• Structured essay-type questions
• Essay-type questions
• Written case studies
• Open-book examination

See chapter 3, section 3.5, for an elaborate discussion of all these measurement instruments.

The following BM III examination papers were taken for analysis:

• Paper I & II of November 1994
• Paper I & II of January/February 1995
• Paper I & II of November 1995
• Paper I of November 1996

An analysis of the exact questions posed in the examination papers produced the following information:

The main action verbs used in the examination papers were the following:

• identify
• explain
• investigate
• give your opinion
• list
• discuss
• develop
• comment on
• criticise
• defend
• provide examples (practical)
• refer to
• describe
• compile a report
• make a recommendation
• make a suggestion

The word “suppose” was frequently used in order to promote some kind of role-play.

In the case of BM III examination questions, it appears as if only a limited range of action verbs were used per examination paper. It was found in the analysis that there are a number of verbs that were used regularly in one or two papers, but not in the rest. Below is a summary of the main findings of the analysis:

• The only action verb that occurred in all seven papers, is: explain. The rest only occurs in some of the papers on an occasional basis.

• The second most common verb appears to be “give your opinion” (in five papers).

• The third most common verb: identify (four papers).

• Examples to be provided in conjunction with the other verbs occurred in five papers.

• A number of verbs that appeared in at least three examination papers are the following:
  - list
  - discuss
  - develop
  - criticise
  - refer to
• The only verb that was used in only two of the examination papers is “make a recommendation”.

• The remaining verbs that occurred in only one of the examination papers, are the following:
  - investigate
  - comment on
  - defend
  - describe
  - compile a report
  - make a suggestion

For the sake of a credible conclusion, the researcher decided only to consider action verbs that were used in at least three of the examination papers. That makes possible the following list of “regular” BM III action verbs:

• explain
• give your opinion on
• identify
• discuss
• develop
• criticise
• refer to
• give examples
• list

The normal cognitive levels that these verbs represent, are as follows:

• explain = comprehension
• give opinion = evaluation
• identify = knowledge
• discuss = comprehension
5.6.3.3 Findings from a comparison between BM III curriculum objectives and measurement instruments

On the basis of the above information, a comparison was drawn between the action verbs used in the objectives of the BM III curriculum and those in the examination questions.

Curriculum objectives action verbs:             Examination questions action verbs

explain                                      explain
discuss                                      discuss
give examples                               give examples
develop                                     develop
give opinion                                give opinion
identify                                    identify
list                                        list
criticise (evaluation level)

implement
manage

distinguish between

monitor                                     criticise
                                          refer to
It seems as if all the action verbs used in the examination questions, could be paired with the action verbs in the objectives of the curriculum. There were however, instances where the exact verbs differed, but on the basis of the cognitive levels, they correlated (for example in the case of criticise vs implement).

It would therefore seem as if a positive correlation also exists between the curriculum objectives of BM III and the measurement instruments used in BM III to determine the realisation of those objectives. The same comments made in the case of BM II, section 5.6.2.3, will also apply in this case.

5.6.3.4 Findings from a comparison of BM III curriculum objectives and measurement instruments with SERTEC prescriptions

Although some of the comments at 5.6.1.4 will also apply in the case of BM III, it should be acknowledged that much clearer efforts were made in BM III to develop application skills among learners. In particular the use of the open-book examination helped to contribute in this respect. However, even the open-book examination suffers from some of the limitations of the written examination (see chapter 3, sections 3.5.1.2 and 3.5.7.2).
5.7 SUMMARY OF FINDINGS

After the analysis of the BM curricula and the examination questions used in BM I, II and III, the following findings were reached:

- The objectives of the BM I curriculum are mainly concerned with fostering a basic knowledge and understanding of the subject.
- The objectives of the BM II curriculum are mainly concerned with fostering a greater knowledge and understanding of BM among learners.
- The objectives of the BM III curriculum are mainly concerned with fostering an adequate understanding of certain subject contents and developing basic application skills in learners.
- There appears to be a positive correlation between the curriculum objectives of BMI, II and III and the examination questions.
- It was also determined that objectives regarding the development of application skills in learners could not really adequately be realised by means of the written examination as a result of its theoretic nature.
- In conclusion, it seems as if the BM curriculum and evaluation methods are still based on a traditional approach, where the main function of evaluation in the curriculum planning process is only to determine whether the objectives of the curriculum were realised (Posner, 1995: 230 - 232; see also section 2.2.4, chapter 2).

5.8 SUMMARY

In this chapter an attempt was made to determine if any correlation exists between the curriculum objectives in BM I, II and III, and the measurement instruments used to
evaluate the realisation of these curriculum objectives. First, the measurement instruments used in this chapter were described. Second, an illustration of the main measurement instruments used by BM faculties or departments at other tertiary educational institutions in South Africa were provided. Following this, a brief summary of measurement instruments used in BM at TSA was given. Thereafter the principles of SERTEC concerning the curriculum in all technikon courses were described. This was followed by an analysis and comparison of the curricula for BM I, II and III with the measurement instruments used on all three levels. Finally, the main findings from these analysis and comparisons were provided.

The next chapter will be devoted to a description of the strategies followed to conduct a number of focus group interviews with learners in BM at TSA. Also an analysis of the papers delivered by a number of distance education scholars will be discussed. The findings of both methods will also be provided.
CHAPTER 6

VIEWS AND PERCEPTIONS OF LEARNERS IN BM AT TSA

6.1 INTRODUCTION

The previous chapter was concerned with the relationship between the BM I, II and III curricula and the measurement instruments used in BM. The main aim was to determine whether the measurement instruments used in BM are indeed aimed at determining to what extent the curriculum objectives were realised. After analysing the curricula of BM I, II and III, the measurement instruments used on all three year levels were also analysed. After comparing the findings in both analyses it was determined that there is indeed a positive correlation between the curriculum objectives in BM I, II and III, and the measurement instruments used. However, it was also found that with respect to the prescriptions set by SERTEC, that there are certain shortcomings in the measurement instruments used in BM.

In this chapter the main focus will be on the views and perceptions of learners in BM at TSA regarding the concept of evaluation/measurement in BM and the possible introduction of alternative instruments. However, the views of a number of South African scholars in distance education will also be reflected. This chapter will provide a full account of the research methodology used, the structure of the research, the findings reached and a summary of the main findings.

6.2 RESEARCH METHODOLOGY

6.2.1 Introduction

As mentioned in chapter 1, section 1.5.2, the main research methodology used in this research, is the qualitative research methodology. In chapter 1 a broad overview of what qualitative research entails, was given. In this chapter merely a description of the main qualitative research techniques used in this research, is given.
The main type of qualitative methodology used in this chapter is the focus group interview. To provide an adequate description of the focus group interview, it is necessary to supply a description of interviews in general and its nature. Focus group interviews are merely one type of interview and to facilitate a proper understanding of the value and usefulness of focus group interviews, it is necessary to provide such a general background.

6.2.2 Interviews

6.2.2.1 What is an interview?

Seidman (1991:2) describes an interview as a “basic mode of inquiry”. He states that the spoken word, in the form of stories, gives a reflection of the consciousness of a human being (1991:1).

Bogdan & Biklen (1992: 96) define an interview as: “a purposeful conversation, usually between two people but sometimes involving more, that is directed by one in order to get information from the other”.

It appears therefore that interviews are instruments used for the purpose of inquiry and obtaining information from other people. The interview is apparently a widely used instrument in qualitative research (Bogdan & Biklen, 1992: 96; Seidman, 1991: 1; Taylor & Bogdan, 1984: 77).

6.2.2.2 The purpose of interviews

Seidman (1991:3) sees the primary purpose of interviews to be an attempt to understand the experience of different people and the meaning such people attach to their experiences. This attempt is normally based on an interest in the “stories” of people and a high regard for those stories. Seidman (1991:4) further states that interviews help a researcher to understand the contexts of the behaviour of people.
6.2.2.3 Advantages of interviews

The following aspects can probably be regarded as some of the main advantages of interviews as a method of inquiry:

- Interviews allow a researcher the opportunity to illuminate very personal and subjective human experiences.

- It allows researchers access to settings or information that are inaccessible by other means.

- It is a very effective method to use when there are time constraints.

- Interviews also allow relative easy access to a broad range of settings and people (Taylor & Bogdan, 1984: 80, 81).

6.2.2.4 Disadvantages of interviews

Seidman (1991: 5-7) mentions the following disadvantages of interviews:

- Interviews can sometimes be a time-consuming and costly process.

- Interviews are a very labour-intensive process.

- Interviews are sometimes regarded by some people to be a form of exploitation of others for the sake of selfish interests (obtaining a degree).

However, the advantages of interviews as a method to get access to information that cannot be obtained by other means, makes it a very useful method to employ.
6.2.2.5 Different types of interviews

Seidman (1991:9) distinguishes two main categories of interviews that are apparently extreme opposites of each other. The first is the very structured and standardised type of interview normally used for doing surveys. The questions used in such interviews are of a very closed-end nature, in other words specific answers are required. The second category consists of the more open-ended and mainly unstructured type of interview.

Seidman (1991:9) however focuses on one specific type of interview which he named the “in-depth, phenomenologically-based” interview. This type of interview belongs to the latter category because the types of questions used are very open-ended in nature. This type of interview aims at letting participants reconstruct their experiences regarding the particular topic at hand.

Marshall & Rossman (1995:80-85) identify the following types of interviews:

- In-depth interviewing: These are interviews conducted mainly in the form of a conversation in which the focus is to get the participant’s own perspectives on a particular topic/subject.

- Ethnographic interviewing: This type of interview aims to identify the cognitive structures that forms the foundation of the world-views of people. It also focuses on gathering the cultural background and contexts of participants.

- Phenomenological interviewing: This a particular type of in-depth interview that is based on the phenomenological approach. Phenomenology as a school of thought, focuses on the experiences of human beings and how people’s world-view is developed by their experiences.

- Elite interviewing: This type of interview is normally conducted with “elite individuals”. These are normally senior and influential people in organisations or
the community and they are normally in possession of the most confidential information and highest expertise in particular fields.

The focus group interview is another type of interview that can be added to this list. However, because it is the central method of inquiry used in this research, it will be discussed separately.

6.2.3 Focus group interviews

6.2.3.1 What are focus groups?

Morgan (1988:9) describes a focus group interview as essentially an interview with a group of participants. The main emphasis in focus group interviews is the discussion and interaction among the participants in the group and the researcher becomes merely a moderator of the group (1988:10).

Krueger (1994:6) describes a focus group as a particular type of group that is distinguished on the basis of its purpose, size, composition and procedures. He also describes it as formerly planned discussions which aim to capture the perceptions of people about a particular topic.

6.2.3.2 The purpose of focus groups

Focus groups can be used either as an independent data collection instrument or as a supplement to other instruments (Morgan, 1988:10).

Morgan (1988:11) lists the following uses of focus groups:

- for orientation of a researcher to a new field;
- to generate hypotheses on the basis of the insights and views of participants;
- for the evaluation of research sites and study populations;
- to develop questionnaires and schedules for interviews;
• to obtain interpretations of the results of former research projects.

Krueger (1994:6,10,11) mentions the following purposes of focus group interviews:
• First, it is normally used to capture the perceptions of people towards a certain topic or subject.

• It can also be used to identify certain human tendencies.

• Focus groups can also be used to encourage self-disclosure by participants.

6.2.3.3 Characteristics of focus group interviews

Krueger (1994:16-21) discusses the following characteristics of focus groups:

• People form a central component of focus groups.

• They are normally conducted in a series format.

• Focus groups are normally homogeneous, but the participants in most cases do not know one another.

• Focus groups can be regarded as an instrument to collect data.

• The nature of data assembled from focus groups, is qualitative.

• Although the questions are normally open-ended, the discussions are focused.

An important recommendation made by Krueger (1994:7) and Morgan (1988:49) is that a preferred format of focus group interviews, is that it should be non-directive in order to ensure that the views and arguments of the respondents are reflected and not those of the researcher.

6.2.3.4 Advantages of focus groups interviews

Morgan (1988:10) see as an advantage of focus groups the fact that it can be used both as a self contained (independent) research instrument or as a supplement to other instruments.
Other advantages identified by Morgan (1988:15-21) are the following:

- Focus groups contain elements of the two main sources for the collection of qualitative data, namely individual interviews and participant observation in groups.
- Focus groups can in fact sometimes yield data that are not easily accessible by means of individual interviews or participant observation.
- A major advantage of focus groups is the fact that it makes possible the observation of participant interaction in a short period of time on a particular topic.
- Focus groups help to place the emphasis on the views of participants instead of those of the researcher.
- Focus group interviews require less formal preparation, for example of questions.
- Focus group interviews can also be regarded as relatively cheaper than individual interviews.
- Focus group interviews are very time-consuming both in terms of collecting large amounts of information and transcribing it.
- Focus group interviews are also relatively easy to conduct.
- Focus group interviews are effective for the exploring of topics and generating hypothesis.

Krueger (1994:34-36) lists the following advantages of focus groups:

- He claims that focus groups is a socially-based process which reflects the social nature of human beings.
- Focus groups allow moderators to probe and by so doing, it helps to explore issues that were not anticipated by the researcher.
- The face-validity of focus group interviews is regarded to be very high.
- Focus groups can be conducted at a relatively low cost.
- The results of focus group interviews are obtained more speedily.
- Focus groups normally ensure a bigger sample size than other forms of qualitative research.
Marshall & Rossman (1995: 84) list the following additional advantages of focus group interviews:

- The use of focus groups is a method that is socially oriented which is a natural feature of human existence.
- It creates the opportunity for the exploration of unanticipated aspects during such interviews and is therefore a very flexible approach.

6.2.3.5 Disadvantages of focus groups

Morgan (1988 16,19-21) mentions the following disadvantages of focus groups:

- Focus groups are subject to certain limitations, for example the fact that it can only rely on verbal behaviour based on interaction within a group and it does not normally take place in a natural setting.
- It is difficult to manage information obtained by means of focus groups in terms of important or less important/relevant information:
- Because focus groups interviews are not based on natural settings, the accuracy of the comments of participants is sometimes not perfect.
- Focus groups furthermore do not automatically reflect individual behaviour.

Krueger (1994:36,37) also lists a number of disadvantages of the focus group interview. These disadvantages are as follows:

- Focus group interviews provide less control to researchers that in the case of individual interviews.
- It is a more difficult process to analyse the data of focus groups.
- Focus groups require experienced and trained interviewers to ensure optimal results.
- There might be a lot of variation between different groups.
- It can sometimes be difficult to assemble group members.
• Focus group interviews also require venues and environments that are conducive to conversation.

Marshall & Rossman (1995:85) also list the following disadvantages of focus group interviews:

• The interviewer can exercise less control in the case of focus groups and time is sometimes wasted on irrelevant aspects.
• Data analysis is normally a difficult process because it is difficult to establish the contexts of participants' comments.
• Focus groups require moderators that are well-trained in the relevant skills.
• They also mention the fact of logistical problems because of the intention to conduct the interview in the form of a conversation.

Despite these disadvantages, scholars in general see focus group interviews as a very useful research instrument. It also appeared to be suitable for the purposes of this research.

The researcher will at this stage describe a supplementary instrument used in this research to support the findings obtained from the focus group interviews.

6.2.4 Seminar

In order to gain a more first-hand and updated reflection of current thinking in SA distance education circles regarding evaluation practices, the researcher decided to organise a seminar on the topic: Evaluation in Distance Education. Participants from the following distance education institutions were invited to deliver papers at the seminar: University of South Africa (UNISA), VISTA University and TSA. The participants who did eventually take part in the seminar were:

• Prof. LJ Van Niekerk - UNISA
• Prof. WJ Frazer - UNISA
The researcher also took part in the seminar, but the paper presented by him was merely a summary of this research up to the date of the seminar.

The main purpose of the seminar was to capture the views of SA scholars on evaluation practices in the changing higher educational environment in SA and in particular in distance education. No specific prescriptions were made in terms of what the speakers had to focus on. The only requirement was that it should focus on evaluation practices in distance education.

The main topics discussed by each of the speakers at the seminar are as follows:

- **Van Niekerk**: He gave a general overview of different evaluation methods currently uses in distance education institutions; the shortcomings of the existing methods and new methods that will have to be introduced in future. He also described a paradigm shift that is currently under way in educational evaluation circles.

- **Frazer**: Frazer's main focus was on the requirement that evaluation methods should be able to generate evidence about the levels of competency learners have achieved.

- **Gericke**: Gericke's main aim was to illustrate and explain the implementation and effects the SOLO taxonomy on evaluation practices. The acronym SOLO stands for: Structure of Observed Learning Outcome.

- **Dresselhaus**: Dresselhaus's main focus was on the use of evaluation methods for learner access purposes.
• Du Plessis: Du Plessis’ main objectives were to describe the evaluation practices at Vista University Distance Education Campus (VUDEC); to indicate existing shortcomings in the system and to suggest ways of improving the system.

On the basis of the above descriptions of the research methodology used in this chapter, the researcher will at this stage describe the processes that were followed.

6.3 STRUCTURE OF RESEARCH

6.3.1 Introduction

The interviewer attempted to conduct the focus group interviews in a non-directive way. This however appeared to be a difficult task initially, because of the following reasons:

• The participants were apparently unfamiliar with this type of interview.
• They did not specifically think about this topic before. They needed time during the interviews to think about it.
• They also did not know one another and initially felt shy or reluctant to speak.

Because of these and other reasons, the interviewer had to pose a number of questions to get the discussions going. A lot of probing had to be done. This initially resulted in a structured and directed format. However as the interviews progressed and the participants became more at ease, discussions developed in a natural way and the interviewer did not have to interfere any longer.

In order to lessen the effect of the initial directive format, the researcher, during the analysis process, only focused on responses of learners where they really made a full statement and not merely answered yes or no to a particular question. The reasoning behind this step is the view that, in order for a learner to make a full statement and even provide reasons for such a statement, some thinking and reflection must have taken place.
It did also happen that some questions that were intended to be open-ended, resulted in a close-ended response (yes or no). The interviewer tried to counteract that by adding "why" to every such question in order to force learners to reveal more about how they felt about a particular aspect.

The interviews were mainly conducted on an ad hoc basis, because attempts to arrange for enough participants to attend the focus group interviews were fruitless. The only interview where learners were notified in advance, was the first one where only two learners arrived for the interview although more than 20 learners were invited. As a result, the researcher decided to contact learners who visit TSA's library and study centre at the Florida campus from time to time. Although the researcher did get hold of an adequate number of learners, they made it clear that they do not visit the campus on a daily basis. It was because of such reasons that the researcher decided to interview the learners on an ad-hoc basis.

However, the ad-hoc format of the interviews also had a number of advantages. Some of the main advantages are the following:

- No pre-determined ideas. The responses from the learners were spontaneous and that gave a relatively accurate picture of what perceptions they had on the topic.

- It also prevented instances of learners answering what they expected the interviewer wanted to hear.

- It was for some a relatively natural setting, because the interviews, with the exception of interviews 1 and 2, were conducted at the study centre where a lot of learners normally congregate on an informal basis.

- It gave a clear picture of how "updated" learners were about an aspect such as evaluation practices. It appeared as if many of them have not yet thought about this
issue critically. They have apparently been so conditioned by the written examination, that they never scrutinized it critically.

As a result of the problem of assembling enough learners together at the main campus, it happened that the number of learners in interview 1 and 6 were not enough to qualify as focus groups. The minimum number of participants appears to be 4 (Morgan, 1988:44). Since the objective with these two groups also was to conduct focus groups interviews with them, the same procedure was followed as in the case of the other four interviews. However, in the analysis process, the researcher used interviews 2 - 5 as the basis for analysis and while 1 and 6 served as a type of comparison factor. No findings or conclusions were based on information that came from interviews 1 or 6 only.

6.3.2 The size and number of focus groups

Seidman (1991:45) suggests two criteria to determine the size and number of interviews to conduct:

- The first criterion is the aspect of “sufficiency”. There should be a sufficient number of participants that will reflect the range of elements or features of a particular population.
- The second criterion is the aspect of “saturation of information”. This refers to the point where an interviewer realises that he/she is no longer hearing anything new or different from new participants, but that the same things are repeated. A reasonable number of participants altogether is 25 (Seidman, 1991:45).

Morgan (1988:43) states that the average size of a focus group normally ranges between 6 and 10 with the average size being 8. The minimum number of participants in a focus group appears to be four and the maximum 12 (1988: 44).
Morgan (1988:42) argues that the average number of focus groups to have is 3 or 4. He also recommends the aspect of "saturation of information" to determine the number of focus group interviews to conduct.

Krueger (1994:78) also describes the average size of a focus group to be about 6-12 participants. He states that the number of focus groups can vary from three up until several dozens thereof (1994:6). He recommends that the ideal size is between 6 and 9 participants.

Marshall & Rossman (1995:84) also agrees that the average number of participants in a focus group interview is 7-10. They state that focus groups are normally conducted over a number of sessions, but every time with a new group of participants. The aim is to capture specific trends in terms of perceptions and views which is done by means of careful analysis.

In this research, a total of six interviews were held. However, because of the number of participants, interview one and six will probably not qualify as focus groups. Each of them only had two participants. However the other interviews had an adequate number of participants as illustrated below:

- Interview 2: 11 participants
- Interview 3: 9 participants
- Interview 4: 11 participants
- Interview 5: 11 participants

Based on the number of focus group interviews suggested by the authors above, and based on the fact that the researcher could feel that the point of saturation of information was reached, the researcher regarded this number of interviews as sufficient. By Wednesday 3 September 1997, during the fourth interview, it became
clear that the point of saturation of information was reached. No more new responses were forthcoming. The aspect that all the participants in all the groups, including interviews 1 and 6, touched on, was the aspect of practical exposure to real job situations and practical forms of evaluation.

6.3.3 Selection of participants (sampling)

Seidman (1991:41, 42) states that random selection of participants as normally done in quantitative research, is not the way of operation in the case of interviews. The process that is in fact used is called "purposeful sampling". Purposeful sampling focuses on particular features of a wider population, for example:

- typical case
- extreme or deviant case
- critical case
- sensitive case
- convenience sampling
- maximum variation sampling

(Seidman, 1991:41, 42)

Seidman (1991:42,43) recommends in particular the principle of maximum variation sampling. This type of sampling refers to a specific attempt to select a range of participants, and sites, that represent the different "elements" of the wider population. In other words, the participants should reflect the different types of people in a particular population.

Krueger (1994:6) states that the composition of participants is normally done on the basis of similar or common characteristics between the members of the group. He states that homogeneity is normally based on aspects such as: occupation, past use of a program or service, educational level, age, gender, education or family characteristics.
Morgan (1988:45) advises that in order to obtain representativity, the participants chosen for focus groups should reflect particular subgroups from the wider population. He further recommends that a qualitative researcher should specifically target subgroups that will provide the most relevant and useful information regarding the topic for discussion. Morgan further stresses the necessity for the researcher to clearly acknowledge a particular subgroup or bias that was used, so that no unjust generalisation is undertaken. The apparent basis upon which groups are selected can be one of the following: gender, race, age and social class (Morgan, 1988:46).

In this research it was decided to use a subgroup approach. The participants consisted of black learners from the Gauteng province in South Africa. Reasons for this selection are the following:

- Cost factors. It would have been a very expensive process to assemble learners from different provinces. The main campus of TSA is located in Florida, Gauteng.
- The learners that took part in the focus group interviews come to the main campus on a regular basis. Most of those learners are black.
- Attempts to assemble enough learners from different racial groups failed.

However, the aspect of similarity also played a determining factor. The researcher made sure that only learners registered for Business Management I, II, and III took part in the interviews. The names and learner numbers of the learners were collected and verified. Attempts were also made to have groups that consisted of first year, second year and third year learners respectively. However, the number of focus groups for each year level were not enough to warrant a meaningful analysis of the views of learners per specific year level. The following composition was achieved:

Interview 1: Third year learners

Interview 2: First year learners

Interview 3: Mixed
Interview 4: Mixed

Interview 5: Mixed

Interview 6: Second year learners

The views portrayed later in this chapter, section 6.5, will therefore be regarded as the views of learners from all three year levels. No group composition was however arranged on a gender basis and all groups, with the exception of group 1, consisted of male and female participants.

6.3.4 Types of questions

An attempt was made in this research to pose mainly open-ended questions and to allow learners to discuss them without much interference (see annexure B). The purpose with such open-ended questions was to let the discussions develop spontaneously so that learners could reveal their true perceptions and experiences about the topic for discussion. The intention was also to limit direction by the researcher so that his influence in the process remained minimal. Initially it was difficult to stimulate discussions without interference by the researcher, because it appeared as if the learners needed a lot of probing and direction. In the course of the interviews it happened frequently that questions invoked a yes/no response. In instances such as that, the researcher did a lot of probing in order to stimulate greater revelations of the views and perceptions of learners. As the interviews progressed, participants started to respond more spontaneously and voluntarily and the researcher only had to act as moderator.

6.3.5 Tape recordings of interviews and transcriptions

Seidman(1991:86,87) expresses himself strongly in favour of tape-recordings of interviews and the transcribing of such interviews. He reiterates the belief that the
words of participants are a reflection of their consciousness. He also stresses the importance of recording the exact words of participants.

Seidman (1991:88) recommends a full recording of the complete interview because he argues that the omitting of sections of the interview might lead to a pre-mature judgement about the importance of the information. Seidman even recommends the recording of non-verbal signals which can include laughing, coughing, sighing and other interruptions.

In this research the researcher also used tape recordings and transcriptions. The permission of the participants was sought in each case. However, the researcher did not follow the strategy of full recording where almost everything was recorded as suggested by Seidman. Only the verbal responses were transcribed. The reason for this approach was the fact that the researcher focused on the perceptions of learners and not as such on the way in which they expressed themselves. It is also important to note that the quotations reflect the exact words of the participants and no language editing was done.

6.3.6 Duration of sessions

Seidman (1991:13) states that there is no fixed time-period for all interviews, but that the recommended time should be approximately 90 minutes per interview. The average duration of the focus group interviews in this research was 60 minutes because the researcher reckoned that he could obtain enough relevant information in the course of one hour. The discussions also remained on track most of the time and therefore little time was wasted on irrelevant aspects.

6.4 ANALYSIS OF DATA FROM FOCUS GROUPS

The analysis process was conducted in the following way:
First, the researcher examined all the transcripts to find some particular categories or topics that received more or less the same emphasis in all the interviews. Although a number of leading questions were initially posed, the greatest part of the interviews was not based on specific questions. The analysis could therefore not be based on a comparison of responses to particular questions.

Second, after distinguishing the main categories, the researcher wrote down some direct quotations of the statements and views of different participants on each category. This approach is based on one of the strategies suggested by Morgan (1988: 70), where exact quotations of the views of participants are provided. However, only the core part of participants' responses were quoted in this research.

Third, the similarities and areas of agreement between these responses were examined. These similarities were then interpreted by the researcher.

6.5 FINDINGS BASED ON THE ANALYSIS OF TRANSCRIPTS

6.5.1 Main categories:

After analysing all the transcripts, a number of categories could be identified by the researcher, for example:

- Open book versus closed-book examinations
- Simulations versus actual performance
- The case study
- Oral examinations
- The different types of questions

However, these categories were not discussed in all four the main groups and the researcher decided not to use them for comparison purposes. The categories that
represented all four main focus group interviews and the two additional ones, were the following:

- The importance of evaluation
- The purpose of evaluation
- Difficulties/problems regarding the written examination
- Written examinations versus practical evaluation

6.5.1.1 Category 1: The importance of evaluation

Discussions regarding this aspect were initiated by one of the opening questions asked by the researcher at the start of all the focus group interviews. The question was:

- "What are your views or perceptions regarding evaluation in education in general?"

On the basis of the responses, the following could be determined:

- It appears as if the majority of participants in the four main focus groups agree on the importance of and the need for evaluation in the course of their studies. See the examples below:

  - "We must write examinations.." (interview 2, p. 1).
  - "I prefer to write examinations" (interview 3, p. 1).
  - "Yes, we write examinations because it is necessary to write an examination" (interview 3, p. 2).
  - "Yea, I think the examination is very important.." (interview 4, p. 1).
  - "Yes, I regard examinations as an essential component of my studies" (interview 5, p. 1).
• This view is also apparently supported by the participants in interviews 1 and 6. Below are some examples of such a common view:

- “I think it is necessary for us to write the final examination” (interview 1, p.1).

- “I regard examinations as important” (interview 6, p. 1).

• In addition to this, when the researcher asked whether there was agreement on this matter, most of the participants expressed their support for these views.

• There was only one dissenting voice in focus group interview 3:

- “I don’t see any importance of writing examinations” (interview 3, p 3).

• Finally, it seems that, with the exception of one, there is general agreement concerning the important role of evaluation/measurement in the education process among the participants.

6.5.1.2 Category 2: The purpose of evaluation

On the basis of the information from the transcripts, the following findings were reached:

It seems as if the two main views held in terms of the purpose of evaluation are the following:

• Evaluation is an instrument to help learners to determine the level or extent to which they understand the work they had to learn. Examples of this view are the following:
“Evaluation is important so that you can prove whether you understand” (interview 2, p.3).

“It ensures that the learner understands what he has learnt” (interview 4, p. 1).

“...it evaluates my understanding of the things I have read previously” (interview 5, p. 1).

“...the examination can show the learner that he/she has some understanding concerning the work” (interview 5, p. 2).

- This view was also expressed during interviews 1 and 6:

“So it (examinations) gives you an opportunity to prove what you have learned, what you understand about the subject” (interview 1, p. 2).

“... so that they can prove they understand what’s been taught to them through the year” (interview 6, p. 2).

- A second view is that the purpose of evaluation is to test learners about what they have learnt. Here are some examples:

“... We must be tested to show what we have learned.” (interview 3, p. 2).

“We have to test ourselves at the end of the year” (interview 4, p. 1).

- This view was also reflected in interview six. Both participants expressed this view, for example:

“The person is to test him- or herself” (interview 6, p. 1).

“...I think the learners need to be tested” (interview 6, p. 2).
• Two additional purposes were mentioned, however only in one interview respectively:

- "At the end of the day you can get your diploma or maybe degree so you can go on in future" (interview 3, p. 1).

- "The examination really prepares you for the future when you are going to do a similar job" (interview 4, p. 2).

6.5.1.3 Category 3: Difficulties/problems regarding the written examination

On the basis of the information in the transcripts, the following findings were reached:

• It appears as if the first and foremost problem experienced by learners is the limited time allowed during the examinations. Here are some examples to substantiate this point:

- "...some learners write slowly, and at the end they are going to fail because they don't complete the paper" (interview 2, p. 5).

- "I need much time when I'm writing. I need to think thoroughly "(interview 3, p. 3).

- "...we have some problems such as finishing in time..." (interview 4, p. 3).

- "Most of the examinations which I have written, I've experienced problems, a time problem" (interview 5, p. 4).

• This problem was also raised during interview one and six:
- "We run out of time when we write these open-book examinations" (interview 1, p. 4).

- "I experienced that I don't read the case study in time" (interview 6, p. 3).

- "Now because it takes some time, you can't even finish your full question.." (interview 6, p. 4).

- The second most common difficulty appears to be the aspect of stress and anxiety. The following quotations help to illustrate the point:

  - "Yea, yea, you are getting full of stress, you know" (interview 2, p. 5).

  - "They get stressed by studying hard" (interview 2, p. 4).

  - "It makes you afraid and scared" (interview 3, p. 7).

  - "When you are not studying well or you are not understanding the subject, you become tense ...you are frightened" (interview 4, p. 4).

- This difficulty was also mentioned in interview 6:

  - "..I experienced anxiety..I experienced fear" (interview 6, p. 3).

  - "It (anxiety) affected me a lot because some of the subjects I couldn't cope with" (interview 6, p. 3).

- A third major point of difficulty appears to be the possible complexity in the nature of certain questions:

  - "Yes, sometimes when you have to write the examinations, you get confused about it" (interview 2, p. 3).
“That is why you find that some learners fail at the end of the year, because they are very confused” (interview 2, p. 3).

“There are sometimes trick questions” (interview 3, p. 8).

This same difficulty is also mentioned in interview 6, for example:

“The most difficult thing is that if the examiner used difficult words in the case study it is hard for the learner to understand..” (interview 6, p. 4).

“Sometimes there is something I don’t understand in the question” (interview 6, p. 4).

The remaining difficulties mentioned are the following:

The absence of real-life/job experiences (interview 1, p. 7; interview 4, p. 6).

The serious consequences of falling ill during examination times (interview 2, p. 3).

Learners merely studying for the sake of passing an examination (interview 2, p. 4; interview 3, p. 11).

The numerous examinations that have to be written in a short space of time (interview 2, p. 4).

Learners studying without real understanding or mastery of skills (interview 3, p. 7; interview 3, p. 11; interview 5, p. 1).

The theoretical nature of written examinations (interview 3, p. 9).
6.5.1.4 Category 4: Written examinations versus practical evaluation

On the basis of the information in the transcripts, the following findings were reached:

This category appears to be the most extensively discussed in all six interviews. Although there are numerous quotations that can support each point, the researcher decided to provide only a maximum of three quotations from every focus group to substantiate every point. There appears to be three distinct views concerning this category. These views are the following:

- A number of learners seem to support the written examination.
- A large number seem to support the introduction of practical evaluation methods.
- A large number of learners support a compromise between written examinations (theory) and practicals (practice).

- The following quotations reflect support for the written examination:

  - "I feel that the written case study adequately portrays the real work situation" (interview 2, p. 6).

  - "I, what I know, is that in Business Management, it is not necessary to do practicals in Business Management. So the written examination is OK" (interview 3, p. 4).

  - "Sometimes they will ask you about the case study and then the case study would be so interesting, you find yourself laughing" (interview 3, p. 9).

  - "It is better to write than to do orals" (interview 4, p. 3).

  - "It is better to write, because sometimes we are shy" (interview 4, p. 3).
- "I can't say if it's fair or unfair, but I think it is impossible to talk together when we are doing a practical. At least in writing we all write together" (interview 4, p. 5).

- "No, I disagree with him. In Business Management we can't perform business where people are watching you, so we have to do the written examination" (interview 5, p. 14).

- "No, I prefer the written one" (interview 5, p. 15).

• Support for the written examination was also expressed in interview 1 and 6, for example:

- "We generally feel happy about the way it (written examinations) is being done" (interview 1, p. 4).

- ".. at the moment when I am writing I have the feeling that I am expressing what I have learned" (interview 1, p. 5).

- "I think the way we are examined is acceptable to me for this present time" (interview 1, p. 7).

- "Yea, we can be tested in writing" (interview 6, p. 2).

- "I think the learner must write. Examination is the best method of evaluation" (interview 6, p. 5).

• The following quotations reflect support for practical forms of evaluation. Some learners initially interpreted practical evaluation to mean learners being given practical exposure to places of employment and so forth. The researcher then provided a brief description of what practical evaluation entails. Following these
explanations, learners still expressed their support for practical forms of evaluation. Below are a few examples of that support:

- “I think the practical examination is good because what you have done with your hands, you cannot forget” (interview 2, p. 7).

- “The learner, he prefers to do practicals rather than reading, because sometimes many learners they take a long time to study” (interview 2, p. 8).

- “At third year level you don’t have to write an examination. You should just go for your practicals” (interview 2, p. 9).

- “But I think it will be better to have practical examinations” (interview 3, p. 6).

- “If you complete your course, the first thing you hear when you look for employment, they say: Do you have any experience of this kind of job? Practicals will help us to get experience” (interview 3, p. 10).

- “I prefer the actual practical, because I know theory, fine. Practical, I know nothing” (interview 3, p. 10).

- “I think we have to do it practically because you have to know what you are doing…” (interview 4, p. 5).

- “Maybe, such as in Business Management, I have to know how I lead the business, how I manage the business.. So I have to go to some companies…” (interview 4, p. 5).

- “I prefer the practical exam” (interview 4, p. 11).

- “I think it is better to apply my knowledge practically where I can see if I have made a mistake or not…” (interview 5, p. 8).
"Most of the people who have passed written examinations only know about theory and they don't know nothing about the actual work" (interview 5, p. 11).

"I prefer doing the work while someone is looking at you, whether you are doing it correctly" (interview 5, p. 14).

- Support for practical forms of evaluation were also expressed in interviews 1 and 6, for example:

- "So the one in which we are there physically, I think is very important" (interview 1, p. 7).

- "I think it (practical evaluation) gives you the real experience and not only theoretically" (interview 1, p. 8).

- "The practical involvement of the learner should be included." (interview 1, p. 8).

- "I mean, if the learner enters the examination room, it should be practical, because even in the companies, what is required is practical work" (interview 6, p. 5).

- "I think demonstration is the best" (interview 6, p. 10).

- "I prefer the actual performance" (interview 6, p. 10).

- The following quotations reflect a willingness from learners to accept a compromise between the use of written and practical examinations. In other words, they have indicated that both a written and a practical examination should be administered:
"The two (practical and written examinations) should go together" (interview 2, p. 2).

"I think they should change this thing of examinations. What I think is that in the first year learners must write an ordinary examination. Then, in their second year, they should write examinations without the text books and in the third year, they should not write any examination but just do practicals" (interview 2, p. 8).

"Yea, you can't just have practicals. You must also have theory in order to know what's going on in business, before you can do practicals" (interview 3, p. 4).

"I think they should go hand in hand. When you have that practical knowledge and you can then write the examination" (interview 3, p. 6).

"Actually I think that practical and theory must go hand in hand" (interview 3, p. 10).

"I think in a written examination, it would be more simple if I have experienced it outside" (interview 4, p. 6).

"I think the best solution is to start with practicals and where you need to become theoretical, you can write" (interview 4, p. 6).

"And when you're writing the examinations, if you have done practicals before it is very easy for you to write, because you will remember what you have done practically" (interview 4, p. 7).

"First of all we must first write an examination, and then after some two years you do practicals" (interview 5, p. 11).
"I say, if I am first year, I must write examination first and second year. At the third year do practicals" (interview 5, p. 11).

"OK, I think both practical and theory." (interview 5, p. 16).

- Support for a compromise between the written and practical evaluation was also expressed in interview 1 and 6, for example:

  - "Theory should not be excluded completely" (interview 1, p. 9).

  - "Case studies should not be done away with. A compromise should be reached between the two (practical and theory)" (interview 1, p. 9).

  - "They (practice and theory) can be reconciled because they work hand in hand" (interview 1, p. 19).

  - "I think the examination, the evaluation method as a whole must be fair, must accommodate both (practical and theory)" (interview 6, p. 5).

  - "So that is why I think practical work is also important. They must accommodate both" (interview 6, p. 8).

  - "So that the learner can be able to get experience and can have theory, to apply theory" (interview 6, p. 8).

6.5.2 Summary of findings from focus group interviews

The findings obtained in the analysis of the transcripts of the focus group interviews can briefly be summarised as follows:
• It was determined that there were four main categories of topics discussed in all the focus group interviews. These categories are the following:

- The importance of evaluation
- The purpose of evaluation
- Difficulties/problems regarding the written examination
- Written examinations versus practical evaluation

• In the first category the following findings were obtained:

- It seems that, with the exception of one participant, there is general agreement concerning the important role of evaluation/measurement in the education process among the participants in all the focus groups.

• In the second category it seems as if the two main views among the participants in terms of the purpose of evaluation are the following:

- First, evaluation is an instrument to help learners determine the level or extent to which they understand the work they have learnt.

- Second, the purpose of evaluation is to test learners on what they have learnt in the course of their studies.

• In the third category the following findings were reached:

- It seems as if the first and foremost problem experienced by learners is the limited time allowed during examinations.

- The second most common difficulty appears to be the aspect of stress and anxiety.
A third major point of difficulty appears to be the possible complexity in the nature of certain questions:

- In category 4 it was found that there seems to be three distinct views in terms of this category. These views are the following:
  - A number of participants seem to support the continued use of the written examination.
  - A large number of participants seem to support the introduction of practical evaluation methods.
  - A large number of participants seem to support a compromise between written examinations and practical evaluation.

At this stage the findings forthcoming from the analysis of the data of the seminar will be considered.

6.6 FINDINGS OF ANALYSIS OF DATA FROM THE SEMINAR

6.6.1 Brief summary of papers

For the purpose of providing a basic description of the main contents of the papers delivered at the seminar, the researcher decided to supply the following summaries of each paper:

6.6.1.1 Van Niekerk

Title: Why evaluate?

Main headings:
• A brief historic overview of perceptions about evaluation
• Knowledge and the curriculum
• Criticism of present curriculum theories
• The social-constructionist curriculum
• The teacher and the curriculum
• The problem of evaluation in distance education
• Evaluation in terms of the old paradigm
• Evaluation in terms of the new paradigm
• The concept evaluation
• Why evaluate at all?
• Criteria for evaluation
• Memoranda
• Evaluation methods
• Self-evaluation techniques
• Essays and short answer tests
• Setting of assignments
• Marks and results

6.6.1.2 Frazer

Title: Evidence-generating strategies in distance teaching when competence is at stake

Main headings:

• A definition of competence
• The components of competence
• The link between competence and performance
• Levels of competence
• The role of skills in competence
• Occupational standards and performance criteria
• The association between the functional analysis and instructional design
• The association between assessment practices and the job description
• Assessing occupational competence in practice
• Generating evidence
• Direct observation of a performance
• Technology on support of the assessment process
• Making judgements
• Implications for the distance education practitioner

5.6.1.3 Du Plessis

Title: Evaluation at VISTA University Distance Education Campus: Today and tomorrow

Main headings:

• What is evaluation?
• VUDEC evaluation structure
• Discussion on the questionnaire
• The main shortcomings of the existing Vista model
• Modes of delivery
• Netech
• What can we strive for in future in terms of learner evaluation at VUDEC?

6.6.1.4 Gericke

Title: Assessing transformative learning: The SOLO taxonomy

Main headings:

• Introduction
• Port-modernism and emancipatory (transformative) learning
• Transformative (emancipatory) learning
• Transformative learning and assessment: a possible methodology for higher distance education
• Need for assessment: critical theoretical perspective
• Quantitative and qualitative learning and assessment
• Assessing distance education learners: the SOLO taxonomy
• Explaining the SOLO taxonomy
• Applying the SOLO taxonomy
• The SOLO taxonomy and distance education

6.6.1.5 Dresselhaus

Title: Access for Success.

Main headings:

• Introduction
• The importance of higher education: Vision and provision
• Purpose of assessment
• Assessment and access: A relevant issue
• Access for success: An important issue
• Redressing imbalances: Push and pull factors
• Access for success: The problem identified
• The management of access
• Access for success: Assessment for improved success
• The extended curriculum
• Success management: Strategy 2

6.6.2. Selection of main categories

The researcher searched for a number of categories that were commonly discussed by most of the speakers, and that specifically dealt with evaluation.
Before the three main categories were selected, the researcher made a list of all possible categories discussed by participants. The names of speakers were added to every topic addressed by them. The following represents the initial list:

- Transformative learning (Gericke, Dresselhaus)
- Older forms/methods of evaluation (Van Niekerk)
- New methods for evaluation/assessment (Van Niekerk, Frazer)
- Shortcomings of existing evaluation models (Van Niekerk, Du Plessis)
- The purpose of evaluation/assessment (Van Niekerk, Gericke, Dresselhaus, Frazer)
- Distance education learners (Gericke)
- Evaluation/assessment in distance education (Van Niekerk, Du Plessis, Gericke)
- The curriculum (extended curriculum) (Dresselhaus, Van Niekerk)
- Competence (Frazer)
- (Directly) observed assessment (Frazer)
- Performance (performance criteria) (Frazer)
- Competence and performance (Frazer)
- Criteria (Frazer, Van Niekerk)
- Assessment in practice (Frazer)
- Evidence (Frazer)
- Future implications/demands (Frazer, Du Plessis, Gericke, Van Niekerk, Dresselhaus)
- Technology in support of assessment (Frazer)
- Evaluation methods (Van Niekerk, Dresselhaus, Frazer, Du Plessis)
- What is evaluation? (Van Niekerk, Du Plessis)

The three main categories decided upon were the ones discussed by at least four of the speakers, namely:

- The purpose of evaluation (4 speakers)
- Different evaluation methods (4 speakers)
• Future implications/requirements for evaluation practices (5 speakers)

6.6.2.1 Category 1: The purpose or need for evaluation in education

After analysing each paper, the following findings were reached:

• Van Niekerk: The main purposes of evaluation mentioned by Van Niekerk (1997: 7) are:
  - To inform educators, learners and possible future employers about learner progress and achievements.
  - It can be used to indicate areas of shortcomings in learners’ understanding of subject matters, their learning methods, their preparation time, their attitudes and motivation towards learning.

• Gericke: Gericke (1997:3) mainly describes the views of certain critical theorists who regard assessment more as a negative and dehumanizing process than a positive one.
  - These theorists deny that the purpose of evaluation is to prove the competence of the persons subjected to it.
  - They regard assessment merely as a repressive control mechanism.

• Dresselhaus: Dresselhaus (1997:4) supplies a list of purposes of assessment, namely:
  - To measure the extent to which specific standards were achieved
  - The monitoring of systems
Quality assurance

The measurement of competencies

To measure learning progress

For selection and admission purposes

For motivation purposes

For the diagnosis of learning problems

For programme evaluation

Frazer: Frazer (1997:8) briefly mentions the following purpose of evaluation:

Evaluation is an instrument used to establish the extent to which particular knowledge, skills, tasks and functions were mastered by learners.

The following conclusions were reached on the basis of the information supplied by Van Niekerk, Gericke, Dresselhaus and Frazer:

First, it becomes apparent that there is a variety of purposes of evaluation

Second, there are certain main purposes that can be distinguished.

Third, it also seems to appear as if there are some schools of thought rejecting the common views about the purpose of evaluation, and who rather regard the role of evaluation as a negative process.
6.6.2.2 Category 2: Evaluation methods

- Van Niekerk (1997:10) describes a wide range of evaluation methods. He first distinguishes the following main general categories of evaluation:
  - observation
  - consideration of parts of the learner's work
  - rating scales
  - check-lists
  - portfolios

- Van Niekerk only provides a description of the various types of evaluation methods without specifically emphasizing any particular one. He does however mention the growing use of self-assessment. He mentions the following types of evaluation:
  - Written as opposed to oral assignments
  - Self evaluation
  - Essay - type tests
  - Short answer tests
    - multiple-choice
    - true - false
    - matching items
    - completion items (1997:10 - 12).

- Dresselhaus (1997:11-12) discusses the following measurement instruments:
  - Diagnostic testing
  - Dynamic testing

- Dresselhaus expresses himself in favour of dynamic testing because according to him, it is inherently focused towards capacity-building and inclusiveness.
• Du Plessis (1997: 6) supplies the following list of evaluation methods that are used at her institution. This information she obtained by means of questionnaires that she circulated at their campus. No mention of preferences is made:

- Self-evaluation
- Practical teaching
- Assignments, examinations
- A modular system
- Projects
- Workbooks
- Portfolios
- Tests for a yearmark

• Frazer (1997: 8-12) discusses various types of measurement instruments. Included in his list are the following:

- Practical demonstrations
- Mental (written or oral) representations

• Frazer distinguishes a specific category of measurement instruments which he calls “evidence gathering methods”. He lists the following types (1997:9):

- observing a candidate
- telling (oral)
- writing
- selection
- evaluation of final product /outcome
- simulations
• The two instruments that Frazer particularly emphasizes are direct observation and competency tests (1997: 10). He argues that direct observation of a learner's performance is one of the best methods of evaluation (1997: 10).

• Frazer also emphasizes simulations as a form of additional strategy to assist in the evaluation process. He mentions the following types of simulations:

  - Case studies
  - Role-playing
  - In-basket tests
  - Models and simulations

• Gericke: Although not specifically discussing different evaluation methods, Gericke does refer to existing evaluation methods such as assignments, examinations and tests, as well as newer forms of evaluation, portfolios and diaries (1997: 6). He actually predicts the future replacement of the first group by the latter.

The following aspects become apparent after the analysis:

• There are quite a variety of measurement instruments available and at the disposal of educational institutions in addition to the written examination.

• Different measurement instruments, other than the written examination, are increasingly being emphasized by scholars.

• It appears as if the emphasis is more on a variety than on single measurement instruments.

6.6.2.3 Category 3: Future implications/requirements

• Frazer (1997: 13-15) mentions a number of implications that the new developments in the educational environment, such as outcomes-based education, competency-
based education, and so forth, will have on distance education institutions. The specific implications with regard to evaluation practices are the following:

- It will demand from lecturers in distance education to know what evaluation methods are most appropriate for determining the extent to which learners have mastered subject contents and skills.

- It will require a reassessment of methods to evaluate learner progress.

- It will require the use of more innovative and less traditional forms of evaluation, for example peer-evaluation, tutor evaluation and the use of subject advisers.

- Distance education institutions will have to demonstrate the methods they use to generate evidence of mastery to bodies such as the National Qualifications Framework.

- More worthwhile and meaningful forms of evaluation will have to be introduced by distance education institutions (Frazer, 1997: 14).

- Du Plessis (1997: 11) makes the following recommendations to her own institution, VUDEC, with regards to future evaluation practices:

  - First, she recommends a greater use of self-evaluation.

  - Second, the use of micro-lessons for the purpose practical evaluation.

  - Third, more opportunities for practical training should be arranged at Vista's residential campuses.

  - Finally, the introduction of a mobile laboratory for departments such as biology and physical science.
• Du Plessis also refers to comments made by the South African Institute for Distance Education (SAIDE), that the written examination is not adequate or suitable for the practical demonstration of skills and knowledge.

• Gericke (1997: 6) envisages the following developments and requirements in terms of evaluation in distance education in the light of new developments in the educational environment:

  - First, he emphasizes a greater use of continuous evaluation in order to facilitate a gradual increase in competence.

  - Second, he recommends that evaluation practices should at all times be reflective of real life and actual job-related tasks.

  - He predicts the introduction of other types of measurement instruments such as portfolios and diaries in the place of assignments, examinations and tests in the context of a transformative learning system.

• Other recommendations made by Gericke are:

  - That the traditional system of awarding marks for student performance be replaced by grading.

  - Evaluation should focus on learning outcomes.

• Van Niekerk (1997: 6) argues that, as a result of the fundamental changes and developments in the higher education sector, for example, the greater focus on the learner instead of the teacher; the greater emphasis on the process of evaluation instead of merely the products thereof and so forth, a total paradigm shift with regard to evaluation practices has become evident. He provides a clear illustration of the old (current) paradigm as opposed to the new one.
The old evaluation paradigm entails the following:

- false starts and missteps are indicators of poor work
- uses same approach to assess all kinds of activities
- assessment designed, developed and mandated externally to meet needs of policy makers
- assesses discreet bits of knowledge
- reinforces science as a static body of disjointed pieces of information
- individual testing dominates
- used to assign marks
- separated from instruction
- fosters convergent instruction
- favours reproductive knowledge
- student achievement assessed by adding scores only
- focus is on management
- assessment is conducted by the teacher

The new paradigm entails the following:

- false starts and missteps are part of active learning
- uses approaches appropriate to the activities
- assessment designed and developed internally to reflect what was taught
- curriculum-specific assessment serves to improve instruction
- assesses what society values
- reinforces science as dynamic inquiry
- co-operative evaluation
- used to improve instruction
- part of instruction: lines blur between assessment and instruction
- fosters divergent instruction
- assesses student progress by multiple methods: portfolios, projects, articles, papers, debates, mind maps
focus is on learning
self-assessment and self-evaluation
(These are direct quotations from Van Niekerk, 1997: 7, 8).

On the basis of the above data, the following summary can be made:

- The South African Higher Education sector is in a stage of transformation and change which has a number of implications, such as new requirements for evaluation methods used in educational institutions.

- That evaluation methods should be used on the basis of appropriateness.

- The existing evaluation methods have to be re-examined in order to determine its continued suitability.

- The use of more innovative and less traditional forms of evaluation is encouraged.

- Evaluation methods used at educational institutions will have to comply with the requirements of bodies such as the NQF.

- More real-life (job-related) and practical forms of evaluation methods should be used.

- A greater emphasis should be placed on learner self-evaluation.

- There should be a greater focus on practical training and evaluation.

- A greater use of continuous forms of evaluation is encouraged.

- A replacement of the mark awarding system to a grading system is recommended.

- Methods should in future focus on the evaluation of outcomes.
• Evaluation practices should reflect the subject contents and skills that were taught to learners.

• Evaluation methods should in fact become part of the instruction process.

• The use of more meaningful forms of evaluation is required.

• The use of multiplicity of evaluation methods is recommended.

6.6.3 Summary of findings from analysis of seminar papers

• It was possible to distinguish three main categories discussed by the majority of speakers. These categories are the following:

  - The purpose of evaluation

  - Different evaluation methods

  - Future implications/requirements for evaluation practices

• In category 1 the following findings were reached on the basis of the information supplied by Van Niekerk, Gericke, Dresselhaus and Frazer:

  - First, it becomes apparent that there is a variety of purposes of evaluation.

  - Second, there are certain main purposes that can be distinguished.

  - Third, it also seems to appear as if there are some schools of thought rejecting the common views about the purpose of evaluation, and who rather regard the role of evaluation as a negative process.
• In category 2 the following aspects become apparent after the analysis:

- There are quite a variety of evaluation methods (measurement instruments) available and at the disposal of educational institutions in addition to the written examination.

- Different types of measurement instruments other than the written examination, are increasingly being emphasized by scholars.

- It appears as if the emphasis is more on a variety than on single types of measurement.

• In category 3 the following findings were obtained. It seems as if the main statements and recommendations made by the speakers at the seminar are the following:

- The South African Higher Education sector is in a stage of transformation and change which has a number of implications, such as new requirements for evaluation methods used in educational institutions.

- That evaluation methods should be used on the basis of appropriateness.

- The existing evaluation methods have to be re-examined in order to determine its continued suitability.

- The use of more innovative and less traditional forms of evaluation is encouraged.

- Evaluation methods used at educational institutions will have to comply with the requirements of bodies such as the NQF.
- More real-life (job-related) and practical forms of evaluation methods should be used.

- A greater emphasis should be placed on learner self-evaluation.

- There should be a greater focus on practical training and evaluation.

- A greater use of continuous forms of evaluation is encouraged.

- A replacement of the mark awarding system to a grading system is recommended.

- Methods should in future focus on the evaluation of outcomes.

- Evaluation practices should reflect the subject contents and skills that were taught to learners.

- Evaluation methods should in fact become part of the instruction process.

- The use of more meaningful forms of evaluation is required.

- The use of multiplicity of evaluation methods is recommended.

On the basis of the above findings, a comparison between the findings emanating from the focus group interviews and the findings from the seminar will be drawn in the next section.

6.7 FINDINGS FROM A COMPARISON BETWEEN CATEGORIES OF INTERVIEWS AND SEMINAR

A comparison was drawn between the categories from the interviews and those of the seminar. The main aim with this comparison was to determine whether there were any
similarities between the two sets of categories. The following similarities were identified:

- First, in terms of the categories itself, the interviews and the seminar have a category in common, that is the category concerned with the purpose of evaluation/examinations.

- Second, it also seems if both groups regard testing of the extent of progress learners have achieved in acquiring new knowledge, understanding and skills, as one of the main purposes of evaluation.

- Third, it became evident that the aspect of more practical forms of evaluation, which was also emphasized by the participants in the focus groups, was also mentioned by a number of speakers at the seminar as one of the recommended measurement instruments for future evaluation purposes.

- A fourth similarity could be found in the learners' view of a compromise between written and practical forms of evaluation and the speakers' recommendation for a multiplicity of evaluation methods. Not one of the groups utterly rejected the written examination, but rather recommended other forms of evaluation in addition to the written examination.

The implications of these similarities between the categories of the focus group interviews and the seminar will be determined and discussed in the last chapter of this research. The next paragraph will provide a brief summary of the contents of this chapter.

6.8 SUMMARY

This chapter started of with an introduction to the main aims of the chapter and a description of the research methodology used. After providing a description of the uses of the focus group interview as a form of qualitative research, the structure of the
actual research process was described. A description was given of the exact procedures and strategies that were followed. Following this, the main findings obtained from the focus group interviews and the seminar were provided. Finally, a comparison was drawn between the interpretations of the categories of the focus group interviews and those of the seminar.

The final chapter of this research will be concerned with the main conclusions drawn from the findings, the main recommendations based on those conclusions and some closing remarks.
SUMMARY, MAIN FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

7.1 SUMMARY OF RESEARCH

The overall aim of this research was to examine the possible need for and introduction of alternative measurement instruments in Business Management at TSA. Below is a brief summary of the different chapters of this research:

Chapter 1 was mainly concerned with the following aspects:
- An introduction to the main theme of this research;
- A description of how the researcher became aware of the problem;
- A statement of the research problem;
- The specific aims and objectives of this research;
- The research methodology that would be used;
- The scope of this research;
- A definition of certain central concepts; and finally
- A summary of the further structure of this research.

Chapter 2 was concerned with the following aspects:
- An overview of the educational evaluation process, including more recent developments;
- The role and place of evaluation in the education process;
- Different evaluation models;
- The role and place of measurement in the evaluation process;
- Matters concerning educational measurement;
- The implications of recent developments in the South African educational context on evaluation practices; and finally
- The influence of current scientific thinking on evaluation.

In chapter 3 the following aspects were dealt with:
• A definition of the concept, conventional summative measurement instruments;
• A description of the most common summative conventional measurement instruments;
• A global contextualization of the most common summative measurement instruments;
• An indication of the basic measurement instruments used in Business Management at tertiary institutions in South Africa;
• The measurement instruments used in Business Management at TSA;
• Finally, a basic scrutiny of the most common conventional summative measurement instruments.

In chapter 4 the following aspects were examined:

• A definition and description of the concept of alternative evaluation;
• The representative role of performance-based evaluation;
• A definition and discussion of various aspects regarding performance-based evaluation;
• An examination of different types of performance-based evaluation;
• Finally, a brief scrutiny of computer-based evaluation.

Chapter 5 was concerned with the following aspects:

• A description of the main measurement instruments used in BM at other tertiary institutions in South Africa;
• A description of the measurement instruments used in BM at TSA;
• An analysis of the BM I, II and III curriculum;
• An analysis of the measurement instruments used in BM I, II, and III;
• A comparison of the BM curriculum and its measurement instruments.

Chapter 6 was devoted to the following:

• To get a clear understanding of the perceptions of learners in Business Management at TSA as well as South African scholars concerning the written examination and evaluation in general.
• The main research methodology used, was the qualitative approach wherein a number of focus group interviews were conducted with learners and a seminar was arranged for distance education scholars.
• An exposition of the findings of this qualitative research.

This chapter is concerned with a description of the main conclusions reached in this research, the recommendations made by the researcher, and some closing remarks.

7.2 A REFLECTION ON THE QUESTIONS POSED IN CHAPTER 1

The questions posed in chapter 1, section 1.3, were answered in the course of this research.

The answer to the overall question, that is whether the written examination can still be regarded as a suitable and adequate method for the evaluation of the academic achievement of learners in Business Management at TSA, was supplied in both chapters 5 and 6.

The answer to the first sub-question, What is the exact role and function of evaluation and measurement in the total educational process?, was provided in chapter 2 where the exact role and function of the evaluation/measurement process was described.

The answer to the second sub-question, What are the shortcomings or disadvantages regarding the existing summative measurement instruments?, was supplied in chapter 3 where an examination of the currently used measurement instruments was launched and such shortcomings were pointed out.

The answer to the third sub-question, What alternative measuring instruments are available and which of them will be best suited for Business Management?, was provided in chapter 4 where a scrutiny of existing or available alternative measurement instruments was undertaken.
The answer to the fourth subquestion, Is there any correlation between the BM curriculum and the measurement instruments used in BM?, was supplied in chapter 5 where an analysis and comparison of the BM curriculum and measurement instruments were provided.

The final sub-question, What are the views of learners in BM at TSA concerning the existing measurement instruments used in BM?, was answered in chapter 6 where, by means of focus group interviews, an attempt was made to determine the perceptions of Business Management learners regarding the written examination.

7.3 SUMMARY OF MAIN FINDINGS

7.3.1 General findings

The main findings obtained in this research are the following:

- In chapter two it was established that evaluation/measurement plays a very crucial role in the education and curriculum planning process, so much so that no effective teaching or learning can take place without it. It was also determined that the process of measurement is one of the three core components of the evaluation process and that no real evaluation can take place unless the process of measurement is utilised.

- In chapter 3 it was determined, on the basis of literature, that there are quite a large number of shortcomings or disadvantages attached to written forms of measurement. However, it was also found that there are a number of advantages that can be credited to written examinations.

- In chapter 4 it was determined that a number of alternative measurement instruments are available. It appeared as if these alternative instruments have a main element in common, namely the aspect of performance. After an examination
of various performance-based measurement instruments, it was discovered that there are both advantages as well as disadvantages that could be attached to these instruments and that not all of them are suitable in all circumstances.

• In chapter 5 it was found that a positive correlation between BM curriculum objectives and the measurement instruments used in BM, does exist. It was also found that, with regard to the evaluation of the practical skills of learners, that the measurement instruments used in BM do have some shortcomings.

• In chapter 6 it was determined that a section of learners in Business Management at TSA apparently expressed a need for more practical forms of evaluation in addition to the written examination. This view is apparently also supported by a certain section of scholars in distance education who recommended the introduction of multiple forms of evaluation into the distance education sector. It was also determined in this chapter that there is no complete rejection of the written examination, but merely a demand for other instruments in addition to the written examination.

In a comparison between the findings reached in chapter 6 and those in the previous chapters, it also became evident that there are a number of positive correlations. The main correlations that could be identified are the following:

• It became evident that, both in literature as portrayed in chapter 2, and the views of learners in Business Management in chapter 6, there is common agreement about the crucial role and function of evaluation in the education process.

• There also seems to be a positive correlation between some of the shortcomings of the written examination as indicated in chapter 3 and those experienced by certain learners in Business Management. Although not all the shortcomings were experienced by learners in Business Management, it became evident that some of them did at one time or the other experience some of those shortcomings of the written examination.
• A positive correlation also seems to exist between the findings in chapter 4 and the views of learners in chapter 6. Some learners expressed themselves strongly in favour of actual practical/performance-based forms of measurement/evaluation. Actual performance measurement apparently forms the core of all the other types of performance-based measurement instruments (see chapter 4, section 4.5.1).

7.3.2 Specific findings from the document review in chapter 5

The more specific findings in chapter 5 are the following:

• The objectives of the BM I curriculum are mainly concerned with fostering a basic knowledge and understanding of the subject.

• The objectives of the BM II curriculum are mainly concerned with fostering a greater knowledge and understanding of BM among learners.

• The objectives of the BM III curriculum are mainly concerned with fostering an adequate understanding of certain subject contents, but also with developing basic application skills among learners.

• There appears to be a positive correlation between the curriculum objectives of BMI, II and III and the examination questions and formats used.

• It was also determined that objectives regarding the development of learners’ application skills could not be realised adequately by means of the written examination as a result of the theoretical nature of the written examination.

• Finally, it seems as if the BM curriculum and evaluation methods are still based on a traditional approach, where the main function of evaluation in the curriculum planning process is only to determine whether the objectives of the curriculum were realised.
7.3.3 Specific findings from focus group interviews and seminar in chapter 6

7.3.3.1 Findings from an analysis of the transcripts of the focus group interviews

The findings obtained in the analysis of the transcripts of the focus group interviews can briefly be summarised as follows:

- It was determined that there were four main categories of topics discussed in all the focus group interviews. These categories were the following:
  - The importance of evaluation
  - The purpose of evaluation
  - Difficulties/problems regarding the written examination
  - Written examinations versus practical evaluation

- In the first category the following findings were obtained:
  - It seems that, with the exception of one participant, there is general agreement concerning the important role of evaluation/measurement in the education process among the participants in all the focus groups.

- In the second category it seems as if the two main views expressed by the participants in terms of the purpose of evaluation are the following:
  - First, evaluation is an instrument to help students to determine the level or extent to which they understand the work they have learnt.
  - Second, the purpose of evaluation is to test learners about the contents of what they have learnt in the course of their studies.

- In the third category the following findings were reached:
It appears as if the first and foremost problem experienced by learners is the limited time allowed during examinations.

The second most common difficulty appears to be the aspect of stress and anxiety.

A third major point of difficulty appears to be the possible complexity in the nature of certain questions.

In category 4 it was found that there seems to be three distinct views in terms of this category. These views are the following:

- A number of participants seem to support the continued use of the written examination.
- A large number seem to support the introduction of practical evaluation methods.
- A large number of participants seem to support a compromise between written examinations and practical forms of evaluation.

7.3.3.2 Findings from an analysis of the seminar papers

On the basis of an analysis of the seminar papers it was possible to distinguish three main categories discussed by the majority of speakers. These categories are the following:

- The purpose of evaluation
- Different evaluation methods
Future implications/requirements for evaluation practices

• In category 1 the following findings were reached:

- First, it becomes apparent that there is a variety of purposes of evaluation.

- Second, there are several main purposes of evaluation in the education process.

- Third, it also seems as if there are some schools of thought rejecting the common views about the purpose of evaluation, and who regard evaluation as a negative process.

• In category 2 the following aspects become apparent after the analysis:

- There are quite a variety of evaluation methods (measurement instruments) available and at the disposal of educational institutions, in addition to the written examination.

- Different types of evaluation methods, other than the written examination, are increasingly being emphasized by scholars.

- It appears as if the emphasis is more on a variety than on single types of evaluation.

• In category 3 the following findings were obtained. It seems as if the main statements and recommendations made by speakers at the seminar are the following:

- That the South African higher education sector is in a stage of transformation and change and these changes have a number of implications for evaluation methods used in educational institutions.
- That evaluation methods should be used on the basis of appropriateness.

- The existing evaluation methods have to be re-examined in order to determine their continued suitability.

- The use of more innovative and less traditional forms of evaluation is encouraged.

- Evaluation methods used at educational institutions will have to comply with the requirements set by bodies such as the NQF.

- More real-life (job-related) and practical forms of evaluation methods should be used.

- A greater emphasis should be placed on learner self-evaluation.

- A greater focus on practical training and evaluation is recommended.

- A greater use of continuous forms of evaluation is recommended.

- A replacement of the mark awarding system to a grading system is recommended.

- Measurement instruments should in future focus on the evaluation of outcomes.

- Evaluation practices should ensure that the subject contents and skills that were taught to learners, were indeed mastered by them.

- Evaluation methods should in fact become part of the instruction process.

- The use of more meaningful forms of evaluation is encouraged.
7.3.3.3 Findings from a comparison between categories of interviews and the seminar

On the basis of a comparison between the categories of the focus group interviews and the seminar, the following similarities were identified:

- First, in terms of the categories itself, the interviews and the seminar have a category in common, that is the category concerned with the purpose of evaluation/examinations.

- Second, it also seems if both groups regard testing of the extent of progress learners have achieved in acquiring new knowledge, understanding and skills, as one of the main purposes of evaluation.

- Third, it became evident that the aspect of more practical forms of evaluation, which was also emphasized by the participants in the focus groups, was also mentioned by a number of speakers at the seminar as one of the recommended measurement instruments for future evaluation purposes.

- A fourth similarity could be found in the learners' view of a compromise between written and practical forms of evaluation and the speakers' recommendation for a multiplicity of evaluation methods. Not one of the groups utterly rejected the written examination, but rather recommended other forms of evaluation in addition to the written examination.

7.4 CONCLUSIONS AND RECOMMENDATIONS

On the basis of the findings obtained in this research and particularly those findings from chapters 5 and 6, the following conclusions were reached by the researcher:
7.4.1 Conclusion 1:

First, it seems as if there is very little doubt (if any) about the crucial role of evaluation in the education process. Evaluation, and by implication, measurement, is an essential component in the teaching/learning context as well as the curriculum planning process. This conclusion is based on the data obtained by means of literature in chapter 2 in which the important role of evaluation is being emphasized and illustrated (see chapter 2, section 2.2.3 and 2.2.4). It is also based on the views expressed by a certain section of learners in Business Management as indicated in chapter 6, section 6.4.2.1.

Recommendation:

In view of the crucial role and function of evaluation in the education process, it is recommended that the Business Management department at TSA should make every effort to obtain and use the most suitable and effective forms of evaluation for the evaluation of learner achievement in Business Management. This recommendation should be followed despite the possible cost implications. This recommendation is made because of the serious consequences the use of inadequate evaluation methods/measurement instruments might have on the future of learners. Learners who might have become experts in a particular field, may be held back because of unsuitable evaluation methods, while those who might not become equally efficient in the workplace are promoted.

7.4.2 Conclusion 2

Second, it is clear that there are indeed certain deep-rooted shortcomings or disadvantages involved in the written examination. This conclusion is based on the fact that such shortcomings are recorded in literature (see chapter 3, section 3.5) and it has also been confirmed by a certain section of learners in Business Management at TSA (see chapter 6, section 6.5.1.3). It became evident from the focus group interviews conducted in chapter 6 that some learners in Business Management at TSA have indeed also experienced some of the common disadvantages or shortcomings of the
written examination. It also became clear from the interviews that it was not merely exceptional cases where such shortcomings were experienced, because this aspect regarding the shortcomings of the written examination, was one of the categories that could be identified during the analysis of data from the interviews (see chapter 5, section 5.4.2.3).

Recommendation:

It is recommended that research be initiated to find ways to eliminate or minimise some of the most serious difficulties regarding the written examination. Two difficulties that should receive urgent attention are the stress and anxiety factor and the problem of limited time. Possible suggestions in this regard would be, for example, to use the preset question and the open-book examination (see chapter 3, section 3.5.7.1 and 3.5.8.1 for the advantages of the two measurement instruments) more regularly for Business Management I, II and III in order to help reduce the stress and anxiety factor. With regards to time limitations, perhaps it should be considered to do away with time limitations in all written examinations and to allow all learners to complete the examination. Certain broad limitations might still be required, for example a maximum of a full day for an examination, but such limitations should not negatively affect learners.

7.4.3 Conclusion 3:

Third, it became clear that there are quite a variety of alternative measurement instruments available in addition to the written examination. This conclusion is based on the numerous alternative measurement instruments discussed in chapter 4, including computer-based measurement, and the different measurement instruments discussed by speakers at the seminar (see chapter 6, section 6.6.2.2).
Recommendation:

In the light of the need for alternative forms of evaluation as expressed in literature and by a certain section of learners at TSA, as well as the availability of various alternative forms of measurement, it is hereby recommended that lecturers in BM at TSA should strongly consider the introduction of additional measurement instruments for the evaluation of the achievement of learners in BM. Although the selection should be based on the principle of appropriateness, as recommended by some speakers at the seminar, particular attention should be given to the needs expressed by a certain section of learners in BM at TSA, that is for practical (actual performance-) based measurement instruments. Learners should in fact be consulted and involved in the evaluation processes and decisions relating to evaluation matters. This will be in accordance with the requirements of legitimacy as discussed in chapter 2, section 2.3.3.3.

7.4.4 Conclusion 4:

Fourth, it seems that, despite the existence of a positive correlation between the BM curriculum objectives and the measurement instruments used in BM, there are certain shortcomings with regard to the extent to which the instruments actually test the realisation of curriculum objectives. Because of such shortcomings it would seem that the measurement instruments cannot adequately determine to what extent all curriculum objectives were realised. This conclusion is based on the fact that the measurement instruments used in BM to measure the application skills of learners, are subject to the limitations of the written word, that is the problem of being theoretical. It is also based on the fact that a certain section of learners in BM at TSA have expressed themselves in favour of actual practical or performance-based measurement instruments, which is an indication of the shortcomings of written examinations.

Recommendation:

It is hereby recommended that alternative measurement instruments be introduced in
BM at TSA that will ensure that the aspect of practical skills are evaluated adequately. Because of the serious consequences of inadequate measurement instruments, this recommendation should be implemented as soon as possible. The BM curriculum should also be adapted to state more clearly what types of measurement instruments should be used for the evaluation of practical skills. Lecturers in BM should investigate the possible introduction of experiential learning, where learners are placed at particular places of employment so that they can gain the practical skills and be evaluated by means of actual performance-based measurement instruments (see chapter 4, section 4.5.1).

7.4.5 Conclusion 5:

Fifth, it appears as if, as a result of its numerous shortcomings, the written examination on its own is no longer regarded as an adequate measurement instrument for summative evaluation of learner achievement in Business Management by a certain section of learners in Business Management at TSA. Consequently, there is a need for additional or alternative forms of measurement. This conclusion is based on the fact that, during the focus group interviews, a number of learners have expressed themselves in favour of the introduction of more practical forms of evaluation/measurement in Business Management at TSA, in addition to the written examination (see chapter 5, section 5.4.2.4). This conclusion can also be regarded as the answer to the overall question of this research (see chapter 1, section 1.3). It also appears as if the written examination is no longer regarded as an adequate summative measurement instrument in distance education by a certain section of scholars in distance education. This conclusion is based on the recommendation by various speakers at the seminar towards the use of multiple forms of evaluation methods in distance education in addition to the written examination (see chapter 5, section 5.5.2.3).

Recommendation:

It is recommended that the final written examination should no longer be the only or
main summative measurement instrument used for the evaluation of learner achievement in Business Management at TSA. Either another form of summative evaluation should be introduced in conjunction to the written examination, or certain appropriate forms of formative/continuous evaluation should be introduced that carries the same weight as the written examination in terms of the final mark awarded to learners. This recommendation should also be implemented in other subjects and courses at TSA as well as other distance education institutions. The reason for this viewpoint is the fact that speakers at the seminar recommended the use of multiple forms of evaluation with regards to distance education institutions and subjects in general.

7.4.6 Conclusion 6:

It seems as if the written examination is still regarded as a useful form of measurement and that there is still a place for it in the evaluation of the achievement of learners in Business Management at TSA. This conclusion is based on the fact that learners who took part in the focus group discussions did not totally reject the written examination as a measurement instrument, but mainly pleaded for the introduction of other measurement instruments in conjunction to the written examination (see chapter 6, section 6.5.1.4). This viewpoint is apparently also supported by scholars who delivered papers at the seminar. They also did not recommend the total abandonment of the written examination, but merely the use of other evaluation methods in addition to the written examination (see chapter 6, section 6.6.2.2 and 6.6.2.3).

Recommendation:

It is recommended that the written examination be retained as a measurement instrument for the evaluation of learner achievement in Business Management at TSA, as well as in other subjects and at other distance education institutions. However, it should only be used for the evaluation of the theoretical components of a subject. This will also contribute towards the validity of the written examination, because if this approach is followed, the written examination will in actual fact be used only to
measure knowledge or skills that can in fact be measured by means of the written examination. The weight attached to the marks of the written examination should be determined on a proportional basis. Lecturers should determine what percentage of theoretical knowledge is measured compared to the practical component.

7.4.7 Conclusion 7

Finally, it seems as if there is also a move towards the use of a variety or multiplicity of evaluation methods in distance education institutions. This conclusion is based on the recommendations regarding future implications and requirements concerning evaluation methods made by speakers at the seminar (see chapter 6, section 6.6.2.3). It is also based on the various alternative measurement instruments that are available (see chapter 4).

Recommendation:

In the final instance, it is recommended that a variety of measurement instruments be made available for Business Management learners at TSA. Whatever methods are selected however, should be used on the basis of appropriateness and suitability. Learners in BM should be involved in this selection process so that the methods decided upon, will be legitimate and generally accepted.

The above conclusions and recommendations conclude this research. At this stage the researcher only wants to make a few recommendations with regard to further research relating to the main theme of this research and a few closing remarks.

7.5 FURTHER AREAS OF RESEARCH

The researcher is aware of the fact that, because of the limited scope of this research, there are various other aspects that could not be addressed. However, those aspects do warrant further research. Examples of such areas are the following:
• The feasibility of various performance-based measurement instruments in distance education and in Business Management in particular

• The introduction and use of a variety of measurement instruments such as portfolios, diaries, work books and so forth

• The effect of other types of measurement/evaluation on the average performance of learners

• The effect of other types of measurement/evaluation on the job-readiness of learners

• The introduction of alternative and a multiplicity of evaluation/measurement instruments in all other subjects or courses offered at TSA as well as other tertiary distance education institutions.

• How to overcome the problem of high costs attached to certain forms of evaluation.

7.6 CLOSING REMARKS

In the course of this research it appeared as if there is great uniformity among learners and scholars concerning the important role and place of evaluation in the teaching/learning process. It also became evident that evaluation is such a crucial element in the teaching/learning process as well as the curriculation process, that no effective teaching/facilitation or learning can occur without it. In the light of this important role and place of evaluation in the education process, the view of the researcher is that no costs should be spared in order to obtain and implement the most adequate, suitable and effective measurement instruments for the evaluation of learner achievement. This should be the approach not only in terms of Business Management learners, or for learners at TSA, but for all learners at all educational institutions in South Africa and even world-wide.
In conclusion, the researcher wants to express his hope that this research might be of benefit to the field of evaluation, not only at TSA, but also at other institutions in the rest of South Africa and even elsewhere.


Frazer WJ & Nieman MM (eds) 1995: The knowledge assessing modes, performance determinants and instructional preferences of first year students at eight major distance educational institutions in South Africa. Pretoria: UNISA.


Pavlich G & Orkin M (eds) 1993: Academic development at South African tertiary institutions. (A comparative case study conducted by CASE for the Ford Foundation.) Johannesburg: CASE.


* 70%
* 80%
* 90%
* 100%

Head of department/subject (full name please): .....................................................

Signature: ..................................................................... Date: ..................................


ANNEXURE B

Questions to participants in the focus group interviews:

It should be noted that these questions were merely used to get the discussions going and to help stimulate the ideas of participants. The questions were not intended to form the main points of discussion.

- What are your views or perceptions regarding evaluation in education in general?
- Why do you feel the progress or achievements of learners should be evaluated or not evaluated at educational institutions?
- Why do you feel satisfied or unsatisfied with the written examination?
- What types of evaluation methods should be used at TSA?
- What types of evaluation methods should be used in BM?
- What recommendations would you make in terms of the improvement of evaluation practices at TSA and in particular in BM?