

**Enhancing the quality of performance assessment in Agriculture in Botswana schools**

By

**Trust Mbako Masole**

Submitted in partial fulfillment of the requirements for the degree of  
PhD: **Assessment and Quality Assurance**

In the Faculty of Education

University of Pretoria

31 March 2011

Supervisor: Prof Sarah Howie, University of Pretoria

## TABLE OF CONTENTS

TABLE OF CONTENTS.....	i
List of Tables .....	vi
List of Figures .....	ix
List of Acronyms .....	x
Declaration of originality .....	xii
Summary .....	xiii
Acknowledgements.....	xv

### CHAPTER ONE

#### AGRICULTURAL PERFORMANCE ASSESSMENT IN SCHOOLS IN BOTSWANA

1.1 INTRODUCTION.....	1
1.2 BACKGROUND TO THE STUDY .....	1
1.3 THE PROBLEM STATEMENT AND RATIONALE.....	5
1.4 DEFINITION OF TERMS.....	7
1.5 THE RESEARCH APPROACH.....	9
1.6 THE AIM AND RESEARCH QUESTIONS.....	10
1.7 SIGNIFICANCE OF THE STUDY .....	12
1.8 OUTLINE OF CHAPTERS .....	13

### CHAPTR TWO

#### THE CONTEXT OF BOTSWANA

2.1 INTRODUCTION.....	15
2.2 DEMOGRAPHY.....	15
2.3 LANDSCAPE AND CLIMATE.....	16
2.4 ECONOMY.....	17
2.5 BOTSWANA'S EDUCATION SYSTEM.....	18
2.5.1 <i>Structure of the Education System</i> .....	18
2.5.2 <i>Management of the Education Sector</i> .....	20
2.5.3 <i>Education and Curricular Reform</i> .....	20
2.6 EXAMINATION OF SENIOR SECONDARY CURRICULUM.....	25
2.7 TEACHING AGRICULTURE IN SENIOR SECONDARY SCHOOLS.....	26
2.8 ASSESSMENT IN AGRICULTURE.....	27
2.9 ASSESSMENT OF PRACTICALS IN AGRICULTURE .....	30
2.10 TEACHER TRAINING .....	32
2.11 CONCLUSION .....	33

## CHAPTER THREE

### LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

3.1	INTRODUCTION.....	34
3.2	THE ORIGINS OF PERFORMANCE ASSESSMENT.....	35
3.3	CONDITIONS FOR PERFORMANCE ASSESSMENT.....	36
3.4	QUALITY ASSURANCE OF PERFORMANCE ASSESSMENT INTERNATIONALLY .....	42
3.5	ISSUES IN PERFORMANCE ASSESSMENT .....	47
3.6	THE CONDUCT OF PERFORMANCE ASSESSMENT IN BOTSWANA.....	54
3.7	VALIDITY AND RELIABILITY OF PERFORMANCE ASSESSMENT INTERNATIONALLY .....	58
3.7.1	<i>Validity</i> .....	58
3.7.2	<i>Reliability</i> .....	63
3.8	CONCEPTUAL FRAMEWORK OF THE STUDY .....	67
3.8.1	<i>System-Level Factors</i> .....	69
3.8.2	<i>School -Level Factors</i> .....	72
3.9	CONCLUSION .....	75

## CHAPTER FOUR

### RESEARCH DESIGN AND METHODS

4.1	INTRODUCTION.....	78
4.2	PARADIGM UNDERLYING THIS STUDY .....	78
4.3	OVERVIEW OF RESEARCH DESIGN.....	81
4.4	RESEARCH DESIGN FOR BASELINE SURVEY: PHASE ONE .....	84
4.4.1	<i>Research design</i> .....	84
4.4.2	<i>Research methods</i> .....	84
4.4.2.1	<i>Sample and Participants</i> .....	84
4.4.2.2	<i>Instrument Development and data collection strategies</i> .....	85
4.4.2.3	<i>Data Collection procedure</i> .....	87
4.4.3	<i>Data Analysis</i> .....	88
4.5	RESEARCH DESIGN FOR THE INTERVENTION STUDY: PHASE TWO ....	89
4.5.1	<i>The nature of design-based research</i> .....	89
4.5.2	<i>Research design</i> .....	94
4.5.3	<i>The research process</i> .....	94
4.5.4	<i>Data collection</i> .....	95
4.5.5	<i>Data Analysis</i> .....	98

4.6	METHODOLOGICAL NORMS .....	98
4.6.1	<i>Dependability of the Results</i> .....	99
4.6.2	<i>Ethical considerations</i> .....	100
4.7	CONCLUSION .....	101

## CHAPTER FIVE

### AGRICULTURE PERFORMANCE ASSESSMENT PRACTICES IN BOTSWANA

5.1	INTRODUCTION .....	102
5.2	BIOGRAPHICAL DATA .....	102
5.2.1	<i>Teachers' age and gender</i> .....	103
5.2.2	<i>Teachers' and school administrators' experience</i> .....	103
5.2.3	<i>Teachers' and school administrators' qualification and training</i> .....	104
5.2.4	<i>Class size</i> .....	107
5.3	PERFORMANCE ASSESSMENT PRACTICES OF TEACHERS .....	108
5.3.1	<i>The mode of assessment</i> .....	108
5.3.2	<i>Learning autonomy</i> .....	111
5.3.3	<i>Assessment for Learning</i> .....	114
5.3.4	<i>Availability of Resources</i> .....	117
5.3.5	<i>Monitoring and Supervision</i> .....	119
5.3.6	<i>Standardisation of marking</i> .....	122
5.3.7	<i>Attitude towards performance assessment</i> .....	125
5.4	DISCUSSION .....	127
5.5	CONCLUSION .....	131
5.6	IMPLICATIONS FOR DESIGN OF INTERVENTION .....	132

## CHAPTER SIX

### DESIGN, DEVELOPMENT AND EVALUATION OF THE FIRST AND SECOND PROTOTYPES

6.1	INTRODUCTION .....	135
6.2	PRODUCT DESIGN SPECIFICATIONS .....	136
6.3	DEVELOPMENT OF THE FIRST PROTOTYPE .....	140
6.3.1	<i>Description of tasks</i> .....	140
6.3.2	<i>Skills equating</i> .....	141

6.3.3	<i>Task Development</i> .....	141
6.4	FORMATIVE EVALUATION OF THE FIRST PROTOTYPE BY EXPERT GROUP .....	159
6.4.1	<i>Research Design</i> .....	159
6.4.2	<i>Participants</i> .....	159
6.4.3	<i>Data collection strategies</i> .....	159
6.5	EXPERTS' VIEWS AND EXPERIENCES WITH THE FIRST PROTOTYPE	160
6.6	CONCLUSION .....	164
6.7	IMPLICATIONS FOR FURTHER DEVELOPMENT .....	165
6.8	DESIGN OF THE SECOND PROTOTYPE - PILOT.....	166
6.9	FORMATIVE EVALUATION OF THE SECOND PROTOTYPE.....	167
6.9.1	<i>Research design</i> .....	167
6.9.2	<i>Participants</i> .....	168
6.9.3	<i>Data collection strategies</i> .....	168
6.10	RESULTS OF THE EVALUATION OF THE SECOND PROTOTYPE.....	171
6.10.1	<i>Lesson Observation</i> .....	171
6.10.2	<i>Standardising marking</i> .....	175
6.10.3	<i>Students' understanding of standardised assessment materials</i> .....	176
6.10.4	<i>Completion of the assessment instrument (Checklist)</i> .....	177
6.10.5	<i>Record keeping</i> .....	180
6.11	CONCLUSION .....	181
6.12	IMPLICATIONS FOR THE SUBSEQUENT DESIGN .....	182

## CHAPTER SEVEN

### DESIGN, DEVELOPMENT AND EVALUATION OF THE THIRD AND FOURTH PROTOTYPES

7.1	INTRODUCTION.....	183
7.2	DESIGN OF THE THIRD PROTOTYPE .....	183
7.3	EVALUATION DESIGN OF THE TRY OUT .....	186
7.3.1	<i>Aim and research question</i> .....	186
7.3.2	<i>Research design</i> .....	186
7.3.3	<i>Participants</i> .....	187
7.3.4	<i>Data Collection Strategies</i> .....	188
7.3.5	<i>Procedure</i> .....	190
7.4	FINDINGS OF THE TRY OUT .....	193

7.4.1	<i>Participants' experiences with the intervention</i> .....	193
7.4.2	<i>Lesson Observations</i> .....	198
7.5	CHARACTERISTICS OF A PRACTICAL QUALITY ASSURANCE SYSTEM .....	203
7.6	CONCLUSION .....	205

## CHAPTER EIGHT

### SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

8.1	INTRODUCTION.....	207
8.2	SUMMARY OF THE STUDY .....	207
8.3	SUMMARY OF THE MAIN FINDINGS .....	208
8.3.1	<i>How is performance assessment currently conducted in Botswana schools?....</i> .....	208
8.3.2	<i>How does the current practice in schools compare with the policy and procedures for performance assessment? .....</i>	209
8.3.3	<i>How does Botswana's experience compares with the international practice? ..</i> .....	210
8.3.4	<i>How can quality assurance processes for performance assessment be developed to ensure valid and reliable marks? .....</i>	211
8.3.5	<i>What are the characteristics of an effective quality assurance system for ensuring valid and reliable performance assessment nationally? .....</i>	211
8.4	REFLECTIONS ON THE CONCEPTUAL FRAMEWORK .....	213
8.5	REFLECTIONS ON THE RESEARCH APPROACH .....	218
8.5.1	<i>Methodological reflections</i> .....	218
8.5.2	<i>Reflection on researcher's role</i> .....	219
8.6	CONCLUSIONS.....	220
8.7	RECOMMENDATIONS .....	225
8.7.1	<i>Policy</i> .....	225
8.7.2	<i>Training and development</i> .....	226
8.7.3	<i>Practice</i> .....	227
8.7.4	<i>Further research</i> .....	228
	REFERENCES .....	229
	APPENDICES .....	258

## List of Tables

Table 2.1:	BGCSE curriculum subject groupings	24
Table 2.2:	Examination format for BGCSE Agriculture	29
Table 2.3:	Brief description of criteria for assessing practical tests	31
Table 3.1:	Comparison between Botswana and international practice on quality assurance processes for performance assessment	57
Table 4.1:	Sample of participants in the study	86
Table 4.2:	Administrators' and teachers' response rate for the questionnaires	89
Table 4.3:	Criteria for high quality intervention	94
Table 4.4:	Schedule of task implementation in schools	104
Table 5.1:	The response rate of respondents	107
Table 5.2:	Proportion of teachers and school administrators trained to conduct performance assessment	107
Table 5.3:	Teachers who neither received training in performance assessment nor related training in assessment (n = 57)	108
Table 5.4:	Frequency of Form four Agriculture class sizes taught by respondents	108
Table 5.5:	Summary of items and factor loadings from principal components analysis with varimax rotation for mode of assessment (N = 57)	111
Table 5.6:	Characteristics of factors for modes of assessment	112
Table 5.7:	Summary of Items and factor loadings from principal components analysis with varimax rotation for learning autonomy (N=57)	114
Table 5.8:	Variance accounted for by the two-factor solution.	115
Table 5.9:	Summary of items and factor loadings from principal components analysis with varimax rotation for assessment for learning (N=57)	117
Table 5.10:	Summary of items and factor loadings from principal components analysis with varimax rotation for availability of resources (N=57)	119
Table 5.11:	Characteristics of factors for availability of resources	120

Table 5.12: Summary of items and factor loadings from principal components analysis with verimax rotation for monitoring and supervision (N=57)	122
Table 5.13: Summary of items and factor loadings from principal components analysis with Verimax rotation for standardisation of marking (N = 57)	125
Table 5.14: Summary of items and factor loadings from principal components analysis with verimax rotation for perception towards performance assessment (N = 57)	127
Table 5.15: Characteristics of factors for perception towards performance assessment	128
Table 6.1: Performance skills and matching performance objectives	146
Table 6.2: Scoring instrument (Checklist) for the task	150
Table 6.3: Scoring instrument (Scale) for the task	152
Table 6.4: Example of summary marksheet for use by teachers	155
Table 6.5: Detailed description of marking criteria for use during field evaluation	156
Table 6.6: Sample of completed summary marksheet	159
Table 6.7: Reliability coefficients for scales of the tasks and assessment instruments	161
Table 6.8: Summary of experts views about task 1	162
Table 6.9: Summary of experts views about task 2	164
Table 6.10: Summary of experts views about task 3	165
Table 6.11: Demographic information of participants	169
Table 6.12: The tasks selected by teachers	170
Table 6.13: The extent of conducting activities by different teachers	174
Table 6.14: Knowledge of assessment displayed by individual teacher	175
Table 6.15: Availability of physical resources in schools to facilitate performance assessment	176



Table 6.16: Students' understanding of assessment practices	178
Table 6.17: An example of scoring by teachers	180
Table 7.1: Example of summary marksheet with brief notes for each criterion for field work	186
Table 7.2: Background information of respondents	189
Table 7.3: The extent to which teachers' embraced assessment for learning	201
Table 7.4: The extent of teachers' knowledge of assessment	202

## List of Figures

Figure 2.1:	Map of Botswana	16
Figure 2.2:	The structure of Botswana's education and training system	19
Figure 3.1:	Factors affecting the validity and reliability of performance assessment marks	70
Figure 4.1:	The Define, Measure, Analyse, Design, Develop, Implement approach of DFSS	83
Figure 4.2:	Research design	84
Figure 4.3:	The cyclic process of design based research of the CASCADE-SEA study	92
Figure 4.4:	Tessmer's layers of formative evaluation	93
Figure 5.1:	Distribution of teachers' age (n = 57)	104
Figure 5.2:	Teachers' teaching experience (n = 57)	105
Figure 5.3:	Teachers' qualifications(n = 57)	106
Figure 6.1:	Skills equating for task 1	141
Figure 6.2:	The overall task showing each step and general criteria	143
Figure 6.3:	Pictorial presentation of the activities for each skill	144
Figure 6.4:	The occurrence of the extent of instructional behaviour	173
Figure 7.1:	Implementation Plan	193
Figure 8.1:	Characteristics and quality processes affecting validity and reliability of performance assessment marks	213

## List of Acronyms

PEO	Principal Education Officer
SPBEA	South Pacific Board for Education Assessment
CRT	Criterion Referenced Testing
PSLE	Primary School Leaving Examinations
JCE	Junior Certificate Examinations
NRT	Norm Referenced Testing
BEC	Botswana Examinations Council
CD&E	Curriculum Development and Evaluation
TT&D	Teacher Training and Development
DSE	Department of Secondary Education
DPRS	Department of Planning, Research and Statistics
ARG	Assessment Reform Group
AFL	Assessment for Learning
MoE&SD	Ministry of Education and Skills Development
MLHA	Ministry of Labour and Home Affairs
MFDP	Ministry of Finance and Development Planning
MTI	Ministry of Trade and Industry
RNPE	Revised National Policy on Education
ERTD	Examinations, Research and Testing Division
CSO	Central Statistics Office
UNESCO	United Nations Educational Scientific and Cultural Organisation
NCE	National Council on Education
CA	Continuous Assessment
HIV	Human Immune Virus
AIDS	Acquired Immunity Deficiency Syndrome
UNICEF	United Nations International Children's Education Fund
CAPA	Creative and Performing Arts
CIE	Cambridge International Examinations
COSC	Cambridge Overseas School certificate

UCLES	University of Cambridge Local Examinations Syndicate
BGCSE	Botswana General Certificate of Secondary Education
CHEA	Council for Higher Education Accreditation
TDTs	Teacher Developed Tasks
CAFs	Common Assessment Frames
SBA	School Based Assessment
DFSS	Design For Six Sigma
DMADDI	Define, Measure, Analyse, Design, Develop, Implement
SPSS	Statistical Packages for Social Sciences
ANOVA	Analysis of Variance
KMO	Kaiser-Meyer-Olkin
DBRC	Design-Based Research Collective
UP	University of Pretoria
TIMSS	Trends in International Mathematics and Science Study
PIRLS	Progress in Reading Literacy Study
PISA	Progress in Student Assessment
SD	Standard Deviation
CN	Condition
CR	Criteria
QAAD	Quality Assessment and Assurance Department

### **Declaration of originality**

I hereby do declare that this thesis, being submitted for the award of the degree of Doctor of Philosophy (PhD) in the University of Pretoria, is my independent work and it has previously not been submitted for a degree or any other examination at this or any other university.

Trust Mbako Masole

March 2011

## Summary

The quality of education in Botswana is not yet up to standard as there has been emphasis on attainment of Universal Basic Education. Quality in education encompasses a number of factors such as the development of the relevant curriculum, improvement of teacher preparation, development of appropriate learning materials, and improving the methods of assessing pupils (Grisay & Mählick, 1991, cited in Kellaghan & Geaney, 2003). The quality of what is going on in the classroom is judged by the processes and outcomes that are defined qualitatively.

Assessment in Agriculture in Botswana senior schools comprises performance assessment and standardised paper-and-pencil tests. Performance assessment contributes only 20% (MoE&SD, 2001.p.6) yet it is allocated more time than paper-and-pencil tests. The aim of the study therefore was to understand and explore the characteristics and quality processes needed in the performance assessment of Agriculture Form Four students to ensure valid and reliable examinations in Botswana.

The study was guided by two research questions. The first research question was: *How valid and reliable are the performance assessment processes in Botswana schools?* This research question sought to understand how performance assessment was conducted in Botswana schools, and how it compared with the international practice. The second research question was: *How can quality assurance processes be developed in order to produce valid and reliable marks for BGCSE Agriculture performance assessment?* The intention was to develop quality processes for performance assessment in the context of Form Four Agriculture in Botswana, to ensure valid and reliable marks for certification.

A design research was employed in this study in which a baseline survey was conducted and based on the outcome, a quality assurance process was designed which included the development of standard tasks and assessment materials. During the baseline survey, teachers and school administrators completed a questionnaire and were also interviewed. Subsequently, prototypes of exemplar materials were developed iteratively in collaboration with practitioners and formatively evaluated. Feedback from evaluation was incorporated into the redesign and development of successive prototypes.

Findings from baseline survey revealed that the conduct of performance assessment in schools was not standardised, primarily due to the absence of assessment policy and procedures to guide its conduct. Implementation of performance assessment was done by teachers who had insufficient training, in large classes with inadequate resources and received very little support from supervisors both internally and externally. Despite all these, insufficient time was allocated for conducting performance assessment, resulting in teachers forming groups most of the time during the conduct of tasks and assigning a single mark for the group based on the quality of the group's product.

However, findings from the intervention study revealed that entrenching quality assurance processes in the system produced valid and reliable performance assessment marks for certification. The characteristics of a quality assurance system for implementation of performance assessment were the presence of an assessment policy; training and accrediting teachers to assess; an efficient internal and external monitoring system; the provision of adequate resources; applying multiple modes of assessment; and multiple rating of the students.

**Key words:** assessment, assessment for learning, authentic assessment, performance assessment, constructivism, pragmatism, validity of assessment, reliability of assessment, formative assessment, and quality assurance.

## Acknowledgements

Let me express my sincere gratitude to my supervisor, Professor Sarah Howie for guiding me throughout the study. Your professional guidance, encouragement when I was close to breaking and calling it quit, and your wisdom enabled me to successfully complete this study. Let me also thank Professor Tjeerd Plomp, for tirelessly guiding me throughout the study. The comments you made were extremely helpful, and have been safely stored for future reference. When I read them, it was like I was seated in front of you listening to you talking.

I would also like to thank Cilla Nel for editing the work wholeheartedly. It is not an easy thing to do. Let me also thank Dr. Vanessa Scherman for the invaluable guidance she rendered whenever Prof Howie was not in. The role you played was crucial.

I cannot forget my family for their continued support. They were like without a husband/father figure in the house. To my wife, thank you very much. You may not know how much it meant to have you by my side. To my kids, I promise that I will be ever home, and pay back for all the absenteeism.

I would like to sincerely thank the Chief Education Officers for giving me permission to conduct my studies in their regions, despite their schools being over researched. This showed that they were committed innovations that can help students to improve their learning. Let me thank all the schools that participated in the study. The school heads were so cooperative and wonderful. At times some schools heads were like co-researchers because they ensured that I got what I wanted. Thank you very much. Let me also thank all Agriculture Senior Teachers in the schools that took part in this study. Your effort surely has been rewarded.

To all Agriculture teachers who participated in this study, your support ensured that this study becomes successful. You never got tired of seeing me every day visiting your classes. Thank you very much. Thank you all the Form Four students who participated in this study. The information you provided was invaluable. This was done for your own



benefit – directly or indirectly. Let me thank all the experts in Agriculture who participated in the evaluation of the prototypes. Your suggestions were extremely useful.

I would have committed an unforgivable crime if I had not acknowledged the assistance of Zelda Snyman, Zanele Lefika, Sibongile and Sibanyoni. Thank you a lot, you played a significant role in the successful completion of this thesis.

I would also like to thank Andrew Graham for tirelessly editing the manuscript to its final state.

## **CHAPTER ONE**

# **AGRICULTURAL PERFORMANCE ASSESSMENT IN SCHOOLS IN BOTSWANA**

### **1.1 INTRODUCTION**

It is important that an education system should strive to provide quality education to the students. Quality of education is not what the students are told by their teachers to do but what they do to create knowledge of their own. To evaluate if students have learnt something, they have to be assessed. Assessment seems to be the most difficult and unpleasant part of the teaching profession. It is not every teacher who can assess and provide quality information needed for making sound policies to enhance learning.

This chapter introduces the study of enhancing quality of Agriculture performance assessment in schools in Botswana. Section 1.2 gives background to the study situating it within the framework of policy reforms. Section 1.3 discusses the problem and rationale leading to the conduct of this study. Section 1.4 outlines definition of terms as they are used in this study. Section 1.5 outlines the research approach, followed by Section 1.6 which gives the aim and research questions for the study. Section 1.7 describes the significance of the study and finally Section 1.8 gives a brief outline of each chapter.

### **1.2 BACKGROUND TO THE STUDY**

The United Nations and Educational Scientific and Cultural Organisation [UNESCO] (2004) declared that the quality of education was declining universally, despite having advocated universal basic education for all school-age going children during the early 1990s. Since then a number of countries committed themselves and have made significant progress in providing education for all (UNESCO, 2004), with Botswana achieving much in terms of access to education, with “Apparent Intake Ratio (AIR) for both six and seven

year olds being more than 100%<sup>1</sup>, which indicates a high degree of access to primary education” (MO&SD, 2003, p.15), taking into consideration that 42.4 million school-aged children in Africa were out of school by 2002 (UNESCO, 2002). However, emphasis on enrolment without provision of sufficient resources to match the large class sizes has resulted in decline in the quality of education.

Recently, emphasis has been directed to quality of education, as evidenced by ratifications of many international conventions, such as The Rights of the Child (United Nations, 2001a), The Dakar Framework for Action (UNESCO, 2000), and the Millennium Development Goals (UN, 2000). Though quality seems to be an elusive concept (Doty 1996), in assessment it is considered to be “the provision of the information of highest validity and optimum reliability suited to a particular purpose and context” (Harlen, 1994, p. 13).

According to Grisay and Mählck (1991) (cited in Kellaghan & Geaney, 2003), quality in education begins with the development of the relevant curriculum, and improvement of teacher preparation and the methods of assessing pupils (p.13). To evaluate if quality education has taken place, assessment of the curriculum is instituted, hence quality of the educational system is measured by student achievement (Kellagan & Greaney, 2003), not by the physical and human resources provided (Pittman, 2003).

The quality of what is going on in the classroom is of greater importance than the number of children who participate in the education process. The notion of merely filling spaces called ‘schools’ with children called ‘learners’, does not address even the quantitative objectives (UNESCO, 2004). UNESCO therefore defined education as being concerned with processes and outcomes that are defined qualitatively. However, this type of education has been elusive, as evidenced by the number of countries lagging behind or declining in achieving quality, including developed ones, (Greaney & Kellagan, 2001; UNESCO, 2004; Walker, 2006).

In order to address the question of quality in education effectively, firstly it has to be realized that education concerns itself not only with cognitive development, but also with

---

<sup>1</sup> Some pupils enrolled at below or over the official school admission age of six

accumulation of particular values, attitudes, and skills. Good quality education should fulfil the acquisition of all these. Secondly, quality instruction has to be accompanied by appropriate quality assessment strategies (Stiggins, 1997), thus assessment is inseparable from the teaching and learning process. The teacher is needed for mentoring, coaching and assessing students while actively engaged in the activities that result in them acquiring knowledge and skills. Stiggins identified five specific standards that quality assessment has to satisfy, one of which alludes to the appropriate assessment format to be used (Stiggins, 1997, p. 167):

A sound assessment examines students' achievement through the use of a method that is capable of reflecting valued targets. We have different kinds of achievement to assess, and as such have to use different kinds of assessment methods to reflect them – select response, essays, performance assessment, structured responses, direct personal communication with students.

The implication here is that quality learning should be learner-centred and formatively assessed. Research has revealed that cooperative learning, which is a learner-centred approach, encourages students' interaction and development of investigative skills (Greenwood & Gaunt, 1994). Since assessment is essentially finding out the worth of what students do, it is logical that they should be assessed as they work either alone or in groups.

The Government of Botswana has since committed itself to providing accessible quality education to all (Government of Botswana, 1994; Ministry of Education & Skills Development [MoE&SD], 2000); Ministry of Finance and Development Planning [MFDP], 1991, 1997, 2003) to mould the child to fit in the participation of future social and economic activities of the country. This was evidenced by two commissions, instituted in 1977 and 1993, which both recommended Continuous Assessment (CA) to be part of a student's final grade. Unfortunately, during the late 1970's, manpower supply was in serious shortage, hence the implementation of the recommendations were not followed as initially intended. For example, the concentration after the first recommendation was on expanding access to primary education so as to acquire strong

foundation in education. Quality was then, though imperative, inadvertently subjected to secondary treatment, given prevalent financial and human resource constraints (Government of Botswana, 1993).

The Government learned from the first National Commission on Education that just providing equitable access to education was a necessary but not sufficient goal. The second commission, which culminated in The Revised National Policy on Education (RNPE) of 1994, clearly indicated government intentions to improve the quality of education as well as assessment through:

- making the curriculum more practical and pre-vocational
- making the learning process more realistic and resembling the world of work
- moving away from a teaching process where the teacher is the provider of knowledge to a learning process that involves students' participation
- introducing continuous assessment at all levels to reduce pressure on pupils and present a more comprehensive assessment of the individual child's capabilities
- monitoring the quality of the education system
- reducing class sizes at primary level, ultimately to 30 pupils, and raising that of senior secondary to 35
- moving from Norm-Referenced Testing to Criterion-Referenced Testing
- reducing the number of unqualified teachers and upgrading the minimum qualification of primary teachers to diploma level (Government of Botswana, 1994).

To emphasise the commitment to quality education provision, the Government of Botswana established the National Council on Education (NCE) to oversee the implementation of the RNPE recommendations. However, though the RNPE advocated provision of quality education and assessment, a major setback was the absence of quality assessment policy for implementing across the education spectrum, including performance assessment, as discussed in the subsequent section.

The policy would provide direction to common conceptualisation of quality in the context of performance assessment, and what constitutes quality assurance processes that make the performance assessment valid and reliable.

### 1.3 THE PROBLEM STATEMENT AND RATIONALE

The quality of an education system needs to be continuously monitored through mechanisms built into the system. The introduction of (CA) was partly intended to monitor the quality of education and partly to reduce pressure associated with one-off terminal examination, culminating in a more comprehensive assessment of the individual child's capabilities. CA is presently limited to those subjects that are practical in nature, and undertaken in the form of performance tasks.

In Agriculture, performance assessment comprise four practical tests (MoE&SD, 2001), which are implemented over the first five terms of the senior secondary programme. However, it is not clear how these practical tests are to be derived, since there is no policy on CA, hence the variation in the conceptualisation of performance assessment tasks. The researcher, during time as an Agriculture Officer, noticed that schools were engaged in tasks with completely different demands and scope. For example, some were deriving the tasks from the same topic, such as *Vegetable Production*, while others treated skills (see Table 6.1) as performance tasks, and yet others regarded an enterprise<sup>2</sup> as a practical test. In addition to the four practical tests conducted, students also do a project in their final year. The project and the practical tests constitute CA in Agriculture.

Practical tests are assessed and scored by the classroom teacher only, while the project is scored by the classroom teacher and then externally moderated. Moderation is carried out at the end of the course and interrogates marks of the final product. The moderator's mark carries more weight than the teacher's (details of the assessment of performance tasks and the project are discussed in Section 2.9). The moderator is brought in to

---

<sup>2</sup> This is a standalone entity which contributes to the gross income of a farm. For example, vegetable production, poultry production, or grain crop production

improve the reliability of performance assessment, but this tends to lower the validity of the assessment process because the moderator has little knowledge of who actually did the work and how the work was done (processes) (Tindal & Haladyna, 2002), and scores the work based solely on the assessment criteria. Performance Assessment is weighed least (20%) of the three papers in Agriculture (See Sections 2.8 and 2.9), despite school-based assessment (SBA) being considered one of the contemporary educational reforms (Airasian & Russell, 2008; Haynes, 2000; McMillan, 2004; Popham, 2005; Stiggins, 1997; van der Merwe, 2000).

Despite the effort to moderate the marks, the conduct of performance assessment is characterized by numerous problems (Baku, 2008; Grimma & Ventura, 2000; Lennox, 2000; Portal, 2000; Ravoice & Pongi, 2000; van den Merwe, 2000; Yadidi & Banda, 2008). During routine spot-checks by the Researcher to verify the conduct of performance assessment, it was discovered that no traceable or retrievable records were kept by schools to justify the marks awarded. Although only a few schools could be visited countrywide due to the shortage of manpower, through triangulation means of information gathering, such as standardisation meetings for the marking of the project, and training workshops, the problem was found to be widespread.

Workshops organized to train teachers in proper conduct of performance assessment did not yield any positive results, with subsequent visits to schools revealing no significant improvements. Inflated marks continued to be submitted to the examining body for summative purposes. It was then suspected that the following were the main causes of improper conduct of performance assessment leading to inauthentic marks:

- lack of standardised tasks for proper performance assessment implementation
- Inadequate training to handle performance assessment
- Lack of resources
- Lack of motivation due to low weighting (20%) of performance assessment
- Large class sizes leading to high workload
- Inadequate supervision and monitoring

- Lack of commitment by school administration (Chong, 2009; Finn et al., 2003; Jones, 2006; Keightley & Coleman, 2002; Mamary, 2007; Maxwell, 2004; Tindal & Haladyna, 2002; Torrance, 1995).

A number of authors have widely documented how performance marks can be validated. These include having a quality assessment policy in place; having a pool of well trained teachers to assess performance tasks; approval of the schools to conduct assessment; internal and external monitoring; development of assessment criteria; involvement of parents and students in assessment; using multiple assessors; reassessment, and collaborative development of standard tasks and assessment materials, (Broadfoot, 1994; Freeman, 1993; Greenwood & Gaunt, 1994; Harlen, 1994; Harry & Schroeder, 2000; McMillan, 2000; Salvia & Ysseldyke, 1998; Stiggins 1997; Tindal & Haladyna, 2002).

The researcher, as an officer working in an assessment environment, could only influence quality assessment through developing standardised tasks and assessment materials in collaboration with teachers. These will be used by the classroom teachers who are strategically positioned to implement performance assessment in a system entrenched with quality assurance.

#### 1.4 DEFINITION OF TERMS

It is necessary here to clarify key terms used in this study:

*Performance assessment* is an all-embracing term used to include products and processes such as portfolios, projects, and experiments (Johnson, Penny & Gordon, 2009; McMillan, 2004). In the context of this study, it means assessment of practicals conducted during the course of the study to enhance learning, using clearly defined criteria (Nitko & Brookhart, 2007). These practicals may range from short activities that take only a few minutes to projects culminating in polished products, in which case a process or product or both are evaluated.



*Portfolio* is one type of performance assessment which is used to demonstrate the student's attainment of learning in practicals (McMillan, 2004; Nitko & Brookhart, 2007; Popham, 2005), such as keeping records of daily transactions during the conduct of practicals. The teacher can then evaluate consciously selected students' records using clearly defined criteria.

*Authentic assessment* determines the degree to which the performance task approximates realism (McMillan, 2004). Assessment of most agricultural activities has the highest authenticity. They involve direct examination of students' ability to use knowledge to perform a task similar to that encountered in the 'real world', for example, preparing a plot or spraying crops with chemicals.

*Product assessment* is made of a completed piece of work performed by students, such as a 'pruned tree' or 'levelled plot'. The product is the end-result of performance or process, and there are situations in Agriculture when assessment of the product is the desirable goal (Gronlund, 2003).

*Process assessment* is made of activities during the performance of a task, and includes assessment of skills and dispositions (Nitko & Brookhart, 2007). In most cases, product and process assessments are carried out to complement each other, or in situations where one cannot be assessed without the other.

*Formative assessment* is the continuous assessment of learning with the main objective of diagnosing students' weaknesses and strengths to institute corrective action. In this study the term is used interchangeably with *assessment for learning* (Airasian & Russell, 2008).

*Quality*, in this study, means conducting performance assessment fitting the intended purpose and context, or conforming to standards (Richard, 1993) of validity and reliability (Harlen, 1994) to promote learning (Greenwood & Gaunt 1994).

*Quality Assurance* is a systematic approach of ensuring quality products and services through entrenching quality in the system, and may involve training teachers on how to conduct performance assessment, use of standard tasks and clear scoring criteria, provision of enough resources, multiple assessment, and accrediting schools to offer

performance assessment (Doherty, 1994; Walklin, 1992). Quality assurance embraces the concept of *Statistical Process Control* (SPC).

*Process control* is a way of ensuring quality by concentrating on the process of production, and by looking at the system as a whole rather than in fragmented parts (Doty, 1996; Richards, 1993), so as to find the process faults (Doty, 1996) and eliminate them (Goetsch & Davis, 1997) before they could affect the end result. Looking holistically at the factors in the teaching-learning process that directly affect quality of assessment, such as methods, equipment or tools, monitoring and supervision, and teachers' skills to assess, infrastructure, administration, educational materials, cohort of learners, policies and documentation, is a process control which employs the strategy of *Six Sigma* to achieve its goals.

*Six Sigma* is a process that dramatically improves efficiency by designing and monitoring everyday activities in ways that minimize waste and resources, and achieves better, faster, and less expensive products (Henderson, 2006; Wild & Ramaswamy, 2008). It focuses on eliminating those factors that might lower the validity and reliability of performance assessment, at the earliest possible occurrence. For example, if inadequacy in training is identified as a contributory factor to low validity and reliability of performance assessment, it should be addressed at the earliest possible opportunity.

## 1.5 THE RESEARCH APPROACH

To understand and explore the characteristics and quality processes needed in the performance assessment of Agriculture in senior secondary schools, the study employed a design research design. Educational design research is "a systematic study of designing, developing and evaluating educational interventions such as programs, teaching-learning strategies, and materials, products and systems" (Plomp 2008, p. 2). The study was conducted in two phases, the first being to conduct a baseline survey to establish the needs and context of the problem, which entailed describing quality assessment practices

and processes that were ongoing, as well as points of views and attitudes that were held by stakeholders in performance assessment (Cohen, Manion & Morrison, 2000).

Based on the findings of the baseline survey, prototypes of quality tasks and assessment materials were then iteratively designed and developed for implementation in the second phase of the study. The prototypes were developed by practitioners at various stages of the design process, adopting a cyclic approach of design, evaluation and revision (Barab & Squire, 2004; Kelly, 2004; Plomp 2008; Van den Akker, Branch, Gustafson, Nieveen & Plomp, 1999). The final prototype was tried in schools and its success was measured by its practicality (utility) in real contexts (Gravemeijer, 2006). The research approach is discussed in detail in Chapter 4.

## 1.6 THE AIM AND RESEARCH QUESTIONS

Botswana has set itself a goal to achieve Universal Basic Education by the year 2016 (Vision 2016). There is significant progress in achieving that as indicated in Section 1.2 that the Apparent Intake Ratio (AIR) for both six- and seven-year olds is more than 100% (MoE&SD, 2003, p. 15). The government has long committed to providing accessible quality education to all (Government of Botswana, 1994; MFDP, 1991; MFDP, 1997; MFDP, 2003). The commissions on education of 1977 and 1993 both recommended the introduction of CA to be an integral component of certification (Government of Botswana, 1977). Continuous Assessment in Agriculture has been implemented for sometime as a response to the recommendations by Commissions on Education of 1977 and 1993. To date, there is no evidence suggesting towards its success with anecdotal evidence indicating problems of validity and reliability of marks.

The aim of the study therefore was to understand and explore the characteristics and quality processes needed in the performance assessment of Agriculture Form Four students to ensure valid and reliable examinations in Botswana. To fully address this aim, the status quo of performance assessment processes in schools was determined through a baseline survey. Subsequently, iterative design, development and evaluation of

prototypes of standard tasks and assessment materials were carried out. Against the above background, the main research questions and sub-questions guiding this study are:

1. *How valid and reliable are the performance assessment processes in Botswana schools?*

The validity and reliability of the marks produced by teachers at school level are a function of the processes and procedures followed both at school-level and system-level. To ascertain this, one has first to understand those processes and procedures in place. This was achieved through a baseline survey directed by the following three sub-questions:

- a) *How is performance assessment currently conducted in Botswana schools?*

It is important first to understand how the system of performance assessment works, to improve its processes. To fully understand and appreciate the concerns and limitations imposed by the current practice, one needs to consult with the practitioners and stakeholders.

- b) *How does the current practice in schools compare with the policy and procedures for performance assessment?*

Any undertaking in education or any sphere of life should be guided by carefully well thought through policies and procedures. One should therefore examine those policies and procedures to be fully convinced that the assessment practices are conducted properly.

- c) *How does Botswana's experience compares with the international practice?*

Due to globalisation, Botswana is compelled to evaluate its education system based on international best practice to provide high standards of education. There is an inevitable paradigm shift from exporting raw materials to human resources, hence Botswana should not be left behind.

2. *How can quality assurance processes be developed in order to produce valid and reliable marks for BGCSE Agriculture performance assessment?*

For the system to produce valid and reliable marks for certification, quality assurance processes and procedures have to be in place. Some of these are system-based and some are school-based. Among the former is the development of standard tasks and assessment materials to guide teachers in their assessment. For these materials to be relevant and useful, they have to involve stakeholders during development. This research question was addressed by an intervention guided by the following sub-questions:

d) *How can quality assurance processes for performance assessment be developed to ensure valid and reliable marks?*

The production of valid and reliable marks is dependent upon embedding quality assurance processes into the system, particularly entrenching it into the doer. For any intervention to be acceptable to the users, they should be part of the developing team. Practitioners should be able to recognise and develop standard tasks and assessment materials for use to improve the acquisition of knowledge and skills by the students.

e) *What are the characteristics of an effective quality assurance system for ensuring valid and reliable performance assessment nationally?*

The iterative development of assessment materials incorporating formative evaluation should ultimately result in characteristics that are peculiar to the situation at hand, which might differ from the standard one. Education should be based on the students' acculturation, as well as their prior knowledge.

## 1.7 SIGNIFICANCE OF THE STUDY

As argued in Section 1.3, there is anecdotal evidence of problems pertaining to the conduct of performance assessment in Agriculture in Botswana schools due to lack of quality assurance processes. In the context of Botswana, performance assessment constitutes CA which the RNPE recommended in 1994 to be incorporated in the certification at senior secondary school level. However, the recommendation was made taking cognisance of the fact that teachers were not well trained to handle CA, hence

priority was given training teachers thoroughly on the conduct of CA before embarking on it. Currently there is no baseline data for teacher level of proficiency to conduct performance assessment in Agriculture, this study is the first empirical study to establish that and design an intervention to ascertain quality assurance measures.

The establishment of quality assurance measures for Agriculture would serve as the basis for developing policy on CA in general as the problems of performance assessment cut across subjects. Currently, there is no policy guiding the conduct of CA increasing the chances of having invalid and unreliable performance assessment marks. The developed policy would furthermore outline other important aspects of performance assessment such as design of curriculum for training institutions particularly their “Assessment Courses” content. There is little done with regards to performance assessment in Botswana and Africa in general.

## 1.8 OUTLINE OF CHAPTERS

This section is intended to give a synopsis of the chapters that follow.

**Chapter One** has introduced the study, including the problem statement, research approaches and questions and clarification of key terms used. **Chapter Two** presents an overview of Botswana’s Education System, in particular the curriculum reforms and how they have affected assessment in general, and Agriculture in particular. The literature is reviewed in **Chapter Three**, revealing a call amongst researchers for performance assessment; guidelines for developing performance assessment; what constitutes quality assessment; development of performance assessment tasks; and associated scoring criteria. It goes on to examine the validity and reliability of performance assessment and presents the study’s conceptual framework. **Chapter Four** presents the adopted research design and explores the approach followed by this design and why it was preferred. **Chapter Five** is a discussion of the findings of the baseline study and their implications for intervention development. The iterative development of the first two prototypes is discussed in **Chapter Six**. **Chapter Seven** discusses the development and evaluation of

the last two prototypes. Lastly, **Chapter Eight** draws conclusions and makes recommendations emanating from the study.

## CHAPTR TWO

### THE CONTEXT OF BOTSWANA

#### 2.1 INTRODUCTION

This chapter presents the context of Botswana, situating the research in terms of its demography outlined in Section 2.2 and landscape and climate discussed in Section 2.3. The economy of the country is discussed in Section 2.4. These have a bearing on the education system of the country. Focus is then placed on the education system of the country in Section 2.5, based on its structure, management, and curriculum reform for pre-primary education, primary education, junior secondary education, and senior secondary education. Section 2.6 examines the examination of senior secondary curriculum, while Section 2.7 confines itself to teaching of agriculture in senior secondary schools. Assessment in Agriculture is delineated in Section 2.8 and Section 2.9 zeroes on the assessment of practicals. Section 2.10 takes a look at the training of Agriculture teachers. The conclusion of the chapter forms Section 2.10.

#### 2.2 DEMOGRAPHY

The population of Botswana has grown significantly from 650,000 in 1973 (Loken, 1973) to approximately 1,756,700 in 2001 (May, 2006; Ministry of Finance and Development Planning (MFDP), 2001 & 2005). During the 2001 population census, the population growth rate was 2.33%, indicating a lower growth rate from 3.5% between 1981 and 1991 (Republic of Botswana, 2009, p. 16). The preliminary results from the 2006 Demographic Survey show a further reduction in growth rate since the 2001 Population and Housing Census, with about 35% of the population being below the age of 15 and 5% above 65 (Ministry of Trade and Industry [MTI], 2008). Life expectancy at birth is estimated to be 39, representing a decline from 55.6 estimated in 2001 (May, 2006; Republic of Botswana, 2009). In 1991, prior to the HIV and AIDS pandemic, life expectancy had increased to 65.3 years (MTI, 2008).



Botswana is a multiethnic and multilingual country, with approximately 23 different ethnic groups speaking approximately 38 different languages (Tlou & Campbell, 1984). The national language is Setswana, while the official languages are Setswana and English, with the latter being the main medium used in government and business offices. The country covers approximately 581,730 square kilometres, with an average population density of three persons per square kilometre (May, 2006; MFDP, 2003). Traditionally a pastoral society with the majority of people living in rural areas, there has been a migration to urban centres by people in search of employment and better lives.

### 2.3 LANDSCAPE AND CLIMATE

Botswana is a landlocked country, as shown in Figure 2.1, sharing borders with Namibia, South Africa, Zimbabwe and Zambia (MFDP, 2003). The Tropic of Capricorn crosses the central part of the country around Mahalapye, signifying the southern latitude over which the sun may be directly overhead, and thus provide a ‘tropical’ and ‘sub-tropical’ climate to the country. The country has a dry, semi-arid climate with temperatures ranging from as low as  $-5^{\circ}\text{C}$  at night to as high as  $43^{\circ}\text{C}$  during the day. Most of the country is covered



by the Kalahari Desert, occupying almost two-thirds of the country. This is home to the indigenous Basarwa (Khoi-San) people. Plain fertile land is found in the eastern part, where most people live. The northern part of the country is a good tourist attraction because of its natural flora and fauna. The Okavango Delta – one of the seven natural wonders of the World is located in this part of the country.

Figure 2.1: Map of Botswana (Source: May, 2006)

The country experiences erratic rainfall with the mean annual rainfall averaging 450 mm, exceeded by moisture loss through evapo-transpiration, with droughts being common.

## 2.4 ECONOMY

Botswana attained independence in 1966 after over 80 years of being a British protectorate (Tlou & Campbell, 1984), and one of the poorest countries in the world. However, in 1972 national income exceeded expenditure for the first time, following the sale of minerals from the newly exploited mines in Selibe-Phikwe, and almost half export earnings from the cattle industry (Loken, 1973). This made it one of the few countries in Africa to have a balanced budget (Loken, 1973), and it is now classified as an upper-middle income country, with most of the population dependent on agriculture for their livelihood (MFDP, 2009). In 1967, a year after independence, one of the world's richest diamond pipes was discovered at Orapa, and in 1982 another one at Jwaneng.

Botswana is home to a variety of minerals, such as copper, nickel, salt, soda ash, coal, gold, and potash (MTI, 2007). Exploration of these and other minerals is ongoing and has recently led to the discovery of large deposits of coal, which is expected to help satisfy the energy needs of the region for the next decade. The rich deposits of minerals have contributed significantly to the growth of the country's economy, and consequently the education sector (May, 2006; MTI, 2007). For example, according to the MFDP (2006), the MoE&SD received 27% of the 2006/2007 recurrent budget and 9% of the development budget.

Botswana's per capita income was Pula (P) 33,000<sup>3</sup> in 2006 (MTI, 2007). Between 1965/1966 and 2005/2006, real Gross Domestic Product (GDP) growth averaged 9% and total Government expenditure had grown to P22.4 billion by 2006/07, and was mainly locally financed (MTI, 2008, p. 14). Financing of the government budgets from foreign grants declined from 51% to less than 2% over the same period (MTI, 2008, p. 16). This

---

<sup>3</sup>£1 is equivalent to P10.00

facilitated the building of foreign exchange reserves, which amounted to around \$12 billion as of the end of November 2008 (MFDP, 2009).

A report by MFDP (2009) suggests that tourism is another important natural resource which is rapidly growing and has recently surpassed minerals in terms of income generation. The country's political stability coupled with its transparent transactions according to Transparency International, earned it the status of the least corrupt country in Africa and 37<sup>th</sup> in the World. This has contributed significantly to the growth of this sector. The World Bank *Ease of Doing Business* Report also ranked the country 48<sup>th</sup> of the 175 countries in terms of relative ease to conduct business. These conditions resulted in a number of industries opening businesses in the country putting high demand on trained labour force. This triggered the government to concentrate on the education system to meet their demands.

## 2.5 BOTSWANA'S EDUCATION SYSTEM

The growth of the economy pushed the demand for skilled labour force. Expatriate labour force was expensive to sustain and the government had the social responsibility to train its own people. With more financial resources from the mining sector and agriculture, there was need to expand the formal education sector. The discussion now focuses on the education system, mainly the structure, management and curricular reform, to understand the place of performance assessment.

### *2.5.1 Structure of the Education System*

Formal education begins with Pre-Primary for children aged 4-5 years. Primary education is for children from 6 to 12 years while secondary education is for children aged 13 – 17. Anybody could enrol for tertiary institutions, through various routes which award degrees up to doctorates, as represented in Figure 2.2.

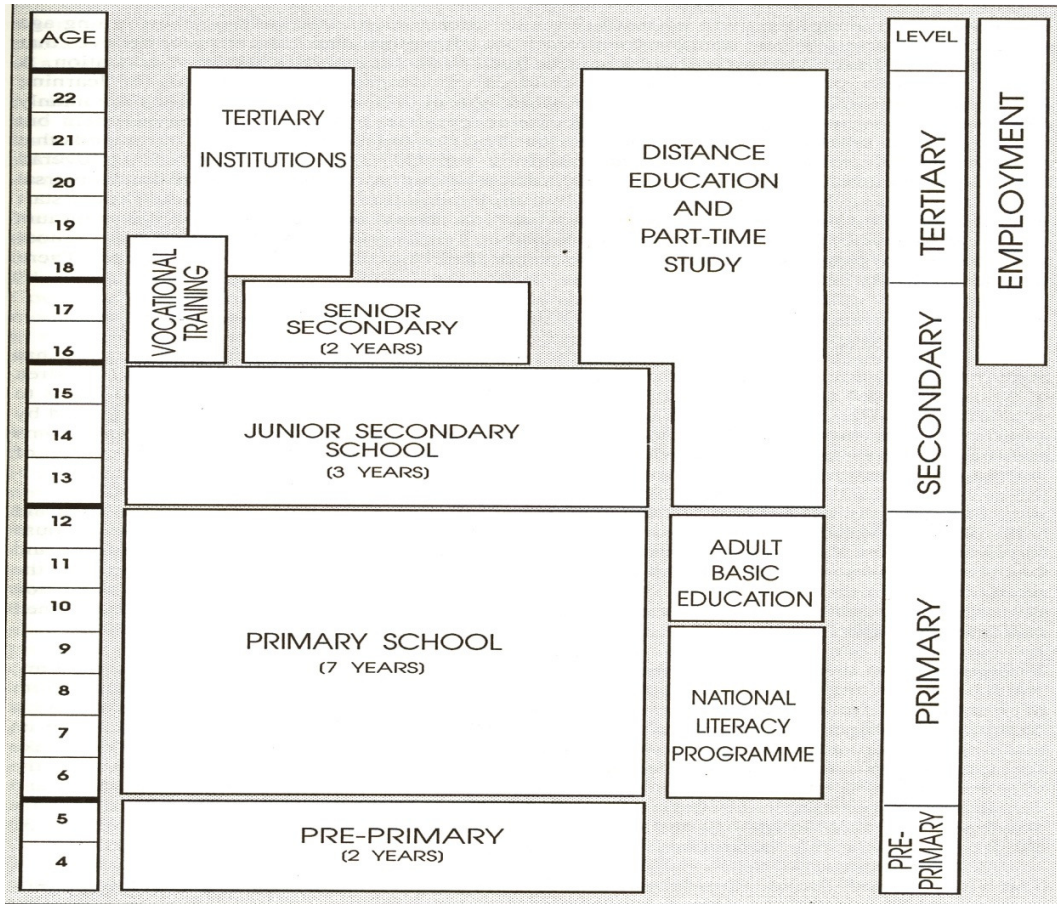


Figure 2.2: The structure of Botswana's education and training system (Source: Republic of Botswana, 1993. p. viii)

The first two years are dedicated to Pre-Primary education, but this has not yet been fully formalised or made operational (See subsection 2.5.3). Primary education runs for seven years, interspersed with examinations at Standard Four and Standard Seven<sup>4</sup>. Running parallel to formal primary schooling is the National Literacy Programme and Adult Basic Education, targeting those individuals who, by circumstances beyond their control, could not be enrolled in the formal education system.

<sup>4</sup> This is equivalent to grade seven

Junior secondary education is a three-year programme, followed by two years of senior secondary schooling and two-to-four<sup>5</sup> years of tertiary schooling (Republic of Botswana, 1993). The structure of the formal education system can be described as 2-7-3-2-(2-4), thus the first 12 years (excluding pre-primary) constitute basic education, in accordance with the World Conference on Education For All in Jomtein (UNICEF, 1990). In addition, the Distance Education run by Botswana College of Distance and Open Learning, University of Botswana and Teacher Training Colleges, offers opportunities to individuals who wish to pursue studies whilst working.

### ***2.5.2 Management of the Education Sector***

Education and Training is mainly the responsibility of the Ministry of Education and Skills Development (MoE&SD), with some ministries such as Ministry of Labour and Home Affairs (MLHA) also offering post-secondary training, and Local Government jointly oversees the running of Pre-Primary education. However, plans are at an advanced stage to wholly relocate this important sector of the education system to the MoE&SD for better coordination, in line with recommendations 9 and 11 of the Revised National Policy on Education of 1994 (Republic of Botswana, 1994).

### ***2.5.3 Education and Curricular Reform***

Reforms have been taking place across all levels of the education system. Accelerated reforms took place particularly at the primary level. There were very little reforms in both the pre-primary and senior secondary levels. Emphasis is now on those two levels.

#### *Pre-Primary Education*

Nearly 10% of children aged between two and five receive Pre-Primary education (MoE&SD, 2009, p. 15). According to MoE&SD (2006, p. 22), only 1,638 children out of a total of 50,868 (3.22%) of Standard One learners had access to Pre-schooling. Currently, there is no common curriculum to link teaching with formal education, and

---

<sup>5</sup> The qualification is dependent upon the institution and programme followed. Normally certificate courses take up to two years, Diploma from two-three years while degree courses are normally four years for pre-service students and two-three years for in-service students

activities vary from one school to the other (MFDP, 1991). The quality of teaching at Pre-Primary is questionable, due to the inadequate training institutions for this level. The only training institution serves the whole country, with an output of only 30 teachers per year. As a result, the number of untrained teachers is high. For example, there were 48.5% in 2005 which increased to 49.6% in 2006 (MoE&SD, 2009, p. 8).

Although progress towards formalising pre-primary education is being made, as evidenced by the establishment of the Pre-Primary Education Unit in the Ministry of education and Skills Development, and the formulation of the relevant policy in 2001 (MoE&SD, 2001), curriculum development is still at a draft stage and substantial Teacher Training has not yet started. No student-teachers have yet enrolled in Colleges of Education for the two-year training programme for pre-primary education.

### *Primary Education*

There were only 250 primary schools at the time of independence in 1966 (MTI, 2008), compared to the latest figure of 782 countrywide (MoE&SD, 2009). During the same period, enrolment rose from 72,000 to 333,417 (MoE&SD, 2009). School fees were abolished in 1980 to facilitate increased access to school by all children, leading to an exponential increase in enrolment, and for the first time including girls, who are now equalling boys in number (Government of Botswana, 2006). It is expected that the quality of teaching has improved since the percentage of untrained teachers dropped from 39% in 1978, to 16% in 1991, to 5.5% in 2009 (MoE&SD, 2009, p. 23).

The proportion of children at school-going age who are not enrolled in schools has fallen, from 17% in 1991 to 3% in 2003 (MTI, 2007, p. 18), with the drop-out rate being only 1.2% in 2006 (MoE&SD, 2009, p. 22). The repetition rate was also low, at 0.2% in 2006 (MoE&SD, 2009, p. 21). This rate, although still undesirable, constitutes an impressive record on the part of government, given that less than two-thirds of children in Sub-Saharan Africa are enrolled in primary schools (United Nations, 2005). Botswana's transition rate from primary to junior school level has been increasing steadily, from 92.6% in 1998 to 97.7% in 2006 (MoE&SD, 2009, p. 21).

In terms of teacher quality, a diploma qualification has been introduced in Teacher Training Colleges to replace certificate qualification and strategies to upgrade in-service teachers to diploma level has been put in place, mainly through distance education. Currently, about 97.1% of primary school teachers have at least a diploma qualification (MoE&SD, 2009, p. 23). The national Pupil-Teacher Ratio is currently 25:1 (MoE&SD, 2009, p. 20), indicating a favourable environment to facilitate pragmatic and constructivist approaches to the teaching-learning process.

The newly developed curriculum has introduced new subjects such as Creative and Performing Arts (CAPA), Guidance and Counselling, and Agriculture, (MoE&SD, 2002, 2005, & 2007). CAPA is made up of different practical subjects such as Design and Technology, Home Economics, Music, Physical Education, Art and Craft, and Business Studies, and was introducing at this level with the aim of developing the manipulative skills in pupils at a young age.

#### *Junior Secondary School Education*

On gaining independence in 1966, there were only nine unified secondary schools in Botswana. By 2008, government had constructed 206 junior secondary schools. In 2006, the transition rate from primary to junior secondary school was 97.7% (MoE&SD, 2009, p. 21), a high figure which resulted in shortage of teachers culminating in 6.6% of expatriate teachers and 1.8% of untrained teachers finding their way in the teaching force (MoE&SD, 2009, p. 28). This triggered massive teacher training initiatives resulting in an oversupply of teachers in all subject areas (Bennel & Molwane, 2008).

The challenge to provide basic education for all resulted in an emphasis on quantity at the expense of quality, (as discussed in Section 1.2), led to the second Education Commission redirecting the philosophy of the education system to providing "... a foundation that enables individuals to cultivate manipulative ability, positive work attitudes..." for its recipients to fit in the world of work (Republic of Botswana, 1993, p. 19). A number of practical subjects were thus introduced into the curriculum, to align with the aim of basic education, in particular. Two such aims which are of relevance here were:

- (i) to include a number of practical subjects that can help learners to develop an understanding and appreciation of technology, manipulative skills and familiarity with tools, equipment and materials
- (ii) to vocationalise the academic subjects.

The present junior secondary curriculum comprises core and optional subjects (MoE&SD, 2002c). The core subjects are studied by all students, who then choose one from each group of Vocational, CAPA and General Studies (MoE&SD, 2002b). Agriculture falls in the core subject grouping, and its assessment is through paper-and-pencil tests as well as a practical work.

#### *Senior Secondary School Education*

There are 28 Government senior secondary schools and a few private schools offering either BGCSE or IGCSE in the country (MoE&SD, 2009). These cannot absorb all students from 206 government junior schools and many more private schools. As such, the current transition rate stands at 67%, but it was expected to have increased to 70% by 2010 when four more new senior secondary schools come into operation (MFDP, 2009, p. 22; MTI, 2008, p. 150).

Assessment at this level of education was localised in 1996, following recommendations by both the First National Commission on Education (Republic of Botswana, 1977) and Second National Commission on Education (MoE&SD, 2002b; Republic of Botswana, 1993). The recommendations of these commissions have resulted in the development of a relevant curriculum to meet the socio-economic needs of the country.

Senior secondary education follows a two-year programme, progressing from the Basic Education Programme. The curriculum is extensive, and offers an opportunity for learners of different abilities to develop their talents. Core subjects are taken by all students, with optional ones from which they choose subjects aligned to their career aspirations. Table 2.1 shows the subject groupings. Subjects are grouped into Core and Optional. Optional group is further divided into Sciences; Creative, Technical and vocational; and Enrichment (2002b). Terminal examinations, which are subject-based,



are written at the end of the programme, and the results are used for selection and placement purposes in training institutions and employment. Only a few candidates progress to tertiary institutions, although this has been steadily increasing.

Table 2.1: *BGCSE curriculum subject groupings*

CORE GROUP	OPTIONAL GROUPS			
Core	Humanities and Social Sciences	Sciences	Creative, Technical and Vocational	Enrichment
English	History	Single Science	Design and Technology	Third Language
Setswana	Geography	Double Science	Agriculture	Physical Education
Mathematics	Social Studies	Chemistry	Art	Music
	Developmental Studies	Physics	Food and Nutrition	Religious Education
	Literature in English	Biology	Computer Studies	Moral Education
		Human and Social	Fashion and Fabrics	Business Studies
			Home Management	

## 2.6 EXAMINATION OF SENIOR SECONDARY CURRICULUM

Before localising examinations in 1996, Botswana students sat for Cambridge Overseas School Certificate (COSC) O-level examinations set and marked by Cambridge International Examinations (CIE), the then University of Cambridge Local Examinations Syndicate (UCLES). The first National Commission on Education (Republic of Botswana, 1977) identified a number of constraints associated with continued dependence on COSC, and these were reiterated by the second Commission on Education:

- Limited ability in influencing curriculum development in line with the aspirations of the nation, in terms of on-going socio-economic development
- The requirement to pass English as a basis for determining the pass levels
- Offering of group subject examinations which aggregate a number of subjects in order to gain a certificate with English determining the grade (Republic of Botswana, 1993, p. 188-189).

Consequently, the notion of establishing an Examinations Council to run the examinations was revisited, with a view to strongly recommend its enactment with immediate effect. This was realised in the National Development Plan 7, 1991-1997 (MFDP, 1991). This was viewed as an opportunity for localising curriculum development to cater for a wider ability group and for emphasizing practical and business subjects. It would also promote continuity and linkages between junior and senior secondary curricula, and allow for the review of modes of assessment so as to relate to the world of work. In 2000, the first groups of subjects were written and marked locally. Currently, almost all the subjects are set and marked locally, with the exception of Religious Education.

## 2.7 TEACHING AGRICULTURE IN SENIOR SECONDARY SCHOOLS

BGCSE Agriculture is classified under *Creative, Technical and Vocational* group (see Table 2.1, above). The *Oxford Advance Learners' Dictionary* (Hornsby, 2000) defines *creative* as “Involving the use of skill and the imagination to produce something new or a work of art”. It defines *technical* as “connected with the practical use of machinery, methods, etc in science and industry”, and *vocational* as “Connected with the skill, knowledge, etc that you need to have in order to do a particular job”. Using these definitions, one may argue that the placement of Agriculture in this group was appropriate, since instruction is largely practical and, consequently, the mode or format of assessment is predominantly performance. A mismatch between instructional strategy and format of assessment can result in wrong data being generated, with serious consequences for those being assessed (Stiggins, 2002). Statute

Agriculture is classified as a ‘Full Classes’ subject according to the MoE&SD Circular<sup>6</sup> of 1<sup>st</sup> February 2005. Full classes are those that take the minimum number of learners as 30, as stipulated in the Revised National Policy on Education. Other *Creative, Technical and Vocational* (See Table 2.1) are classified as ‘Non-Full Classes’ with a maximum of 20 students. Contrary to Subject groupings by CD&E (see subsection 2.5.3), Agriculture is the only subject that has been classified by the Circular as a non-practical subject among the *Creative, Technical and Vocational Group*. It is allocated four periods of 40 minutes per six-day timetable for learning and teaching of theory, conducting practicals and assessment (MoE&SD, 2002b). Teachers are required to have a minimum of five classes (24 periods) and a maximum of six classes (29 periods). Such high workloads given that it involves conduct of performance assessment are likely to impact negatively on achieving the aims of BGCSE Agriculture (MoE&SD, 2000a, p. 2), which are to acquire and develop:

1. an appreciation of agriculture as an applied science.

---

<sup>6</sup> These are directives that are issued by the Ministry Officials as the need arises to modify or direct educational transactions.

2. interest and awareness of existing problems and opportunities in Agriculture in the context of rural development.
3. exposure to out-of-school farming activities, such as agricultural fairs, field trips and the job-shadowing exercise in preparation for the world of work.
4. skills to demonstrate the value of agriculture to the family, community and the national world economies.
5. initiative, problem-solving abilities and scientific methods so as to encourage a spirit of resourcefulness and self-reliance.
6. desirable behavioural pattern and frame of mind in interacting with the environment in a manner that is protective, preserving and nurturing.
7. business and entrepreneurial skills necessary to develop and manage an agricultural project.
8. skills that are relevant to agriculture, including objectivity, precision, initiative, experimentation and research.
9. knowledge and understanding about the efficient use of available government assistance programmes aimed at agricultural development in Botswana.
10. knowledge and understanding of the recent technological development in agriculture.

The aims of BGCSE Agriculture, thus, suggest a paradigm shift from learner-centred approaches to pragmatic and constructivism approaches for effective instruction.

## 2.8 ASSESSMENT IN AGRICULTURE

The examinations conducted by the Botswana Examinations Council are geared towards meeting the overall objective of national education, as pronounced by RNPE, which is “to

assume more effective control of the examination mechanism in order to ensure that the broad objective of the curriculum are realized” (Republic of Botswana 1994,p.5). There are three main assessment objectives for agriculture enshrined in both the Teaching Syllabus (MoE&SD, 2000a) and Assessment Syllabus (MoE&SD, 2000b, p. 4). These are:

1. Knowledge with understanding
2. Handling information, Application and Problem Solving
3. Practical and Investigative Skill

Details of each assessment objective are given in Boxes 1, 2 and 3 (below).

Assessment objective 1 contributes 30% to the examination and mainly constitutes *Knowledge with Understanding*. It is important that students have a basic understanding of the subject matter that will form the fundamental basis for further comprehension of high-order content. Assessment objective 2 assesses the *Comprehension and Critical thinking* and this contributes 40% to the overall mark. Assessment objective 3 assesses the *Practical and Investigative Skills* of the students and constitutes 30% of the overall mark (MoE&SD, 2001, p. 4). Combinations and permutations of the three assessment objectives result in the three papers as shown in Table 2.2 (below), namely; multiple choice (Paper 1), short answer questions and essays (Paper 2), and practicals (Paper 3).

Box 1: Objective 1 - Knowledge with understanding and problem solving

The candidates should be able to use oral, written, symbolic, graphic, tabular, diagrammatical and numerical presentations to:

1. Locate, select, organize and present information from a variety of sources.
2. Translate information from one source to another.
3. Use information to identify patterns, report trends, draw inferences, make predications and propose hypothesis.
4. Present reasoned explanations from phenomenon, pattern and relationships.
5. Solve problems of a quantitative and qualitative nature.

Box 2: Objective 2 - Handling information, application

The candidates should be able to demonstrate:

1. Correct use of terms, symbols, quantities and units of measurement.
2. Correct reference to facts, concepts, laws and principles.
3. Safe Agricultural practices that prepare students for a productive life.

Box 3: Objective 3 - Practical and investigative skills

**Practical Skills and techniques**

The candidates should be able to:

1. Understand and follow instructions.
2. Choose and use suitable techniques, equipment and materials safely and correctly.
3. Record observations, measurements and estimates.

**Practical Investigations**

The candidates should be able to:

1. Identify problem and plan an investigation.
2. Organize and carry out an investigation.
3. Interpret and evaluate observations and experimental data.
4. Draw conclusions and make recommendations.

*(Source: Ministry of Education and Skills Development, Agriculture Assessment syllabus 2001, p. 3-4.)*

Table 2.2: *Examination format for BGCSE Agriculture*

<b>Paper</b>	<b>Nature of the paper</b>	<b>Objective</b>	<b>Duration</b>	<b>Raw mark Weight</b>	
Paper 1	Multiple choice	1 & 2	45 min	40	40%
Paper 2	Short answer questions and essays	1&2	2 hr 15 min	100	40%
Paper 3	Practical	2&3	5 terms*	155	20%

\* 1 term is roughly 66 days  $\pm$ 1, assessment starts already in previous year.

(Source: Ministry of Education and Skills Development, *Agriculture Assessment syllabus 2001*, p. 6.)

The weighting of Paper 1 and Paper 2 is 40% each, whilst that of Paper 3 is only 20% (MoE&SD, 2001, p. 6). The weightings of the papers do not correspond to their demands as evidenced by the time spent. Primarily, a student who performs well in Paper 1 stands a better chance of obtaining a better grade than one who has high marks in Paper 3, yet Paper 3 is allocated more time than any other paper. The pros and cons of multiple choice questions are fully documented by Airasian (2005); Gronlund (2003); Kellagan and Greaney (2001); Nitko and Brookhart (2007) and shall not be discussed here. Grade descriptors presented in Appendix 2.1 attest to the importance of practical skills acquisition. If the practical skills are so important to the learner they should be reflected in the weighting of marks.

## 2.9 ASSESSMENT OF PRACTICALS IN AGRICULTURE

The BGCSE Agriculture practical assessment is divided into two parts. The first part comprises a number of practical tests assessed by the classroom teacher and his/her mark is final. The assessment is guided by marking criteria (See Appendix 2.2), and this part accounts for 51.6% (80 out of 155) of the total mark (MoE&SD, 2001, p. 26). The other part is the project work, which involves problem investigation to design a practical

solution to a real agricultural problem and produce a report on the findings. This accounts for the remaining 49.4% (75 out of 155) (MoE&SD, 2001, p. 27). The total of the two are then scaled down to 20% of the final mark (see Section 2.6) (MoE&SD, 2001, p. 6).

The main aim of the practical tasks assessment is mainly to assess the processes and procedures leading to the outcome, product or artefact. Since the processes have transitory evidence they can only be assessed by the classroom teacher, hence any attempt to moderate the teacher marks will be extremely difficult and distort the outcome. When conducting the practicals a portfolio is kept by the students, detailing the development of the investigation. Examples of tasks for practical work are presented in Appendix 2.2, while the criteria for assessing practical tests are briefly delineated in Table 2.2, (a full guide is presented in Appendix 2.3).

*Table 2.3: Brief description of criteria for assessing practical tests*

<b>Criteria</b>	<b>Description of criteria</b>
Responsibility	the ability to resume responsibility for the task in hand, and to work from given instructions without detailed supervision and help
Initiative	the ability to cope with problems arising on connection with the task, to see what needs to be done and take corrective action.
Technique	the ability to take practical tasks in a methodical, systematic way and to handle tools skilfully and to good effect.
Perseverance	the ability to see a task through to a successful conclusion with determination and sustained effort.
Quality	the ability to attend to detail so that the work done is well finished and well presented.

The objective of the project is to equip candidates with research and investigative skills. It provides students with the opportunity to develop a hypothesis, plan an investigation around the hypothesis, carry out the investigation, analyze and interpret data collected during the investigation, make observations, write a report, draw conclusions and make recommendations.



The time spent on the project depends on its nature. Some take a few weeks, such as surveys, while others last for some months, such as field experiments. The practical processes of carrying out the investigation are not assessed, even though the project is supervised by the teacher. The student then writes a report, which is first scored by the classroom teacher and then externally moderated by a visiting moderator who then reconciles the marks with the teacher. Detailed marking criteria for scoring the project is shown in Appendix 2.4.

Moderators have a final say, despite the model they follow which aims for reconciliation rather than the exertion of external power. Moderators exert power because they are involved with a product-product continuum, with comparability of pupils central to their concern (Radnor & Shaw, 1995). Judging by the examples of practical tasks suggested, and the corresponding marking guide, there is no doubt that the intention is to impart students with life skills and foster critical thinking among students. The determination of whether the practical tasks are carried out and assessed as enshrined in the syllabus is a main aim of this investigation.

## 2.10 TEACHER TRAINING

Teachers for Agriculture are trained at either the College of Education or the College of Agriculture. The former runs a three-year programme for Agriculture teachers destined to teach at junior secondary school level and culminating in a Diploma qualification. The latter trains teachers for degree level and these teachers are meant to teach at senior secondary schools. The course is four years for pre-service students, and three years for diploma holders. A diploma qualification holder can also teach in senior schools and vice-versa. Analysis of the content of “Assessment/Measurement Course” for both colleges revealed that assessment of practicals was not treated in great detail.

One of the recommendations by the Second Commission on Education was the inclusion of continuous assessment marks in the certification of the candidates. This implied that adequate training of teachers to handle continuous assessment should be undertaken. A

consultancy engaged by the examining body to advise on the role and modalities of incorporating Continuous Assessment recommended that the Examining Body should assist the MoE&SD to develop a standard Assessment Course to be taught in Education Colleges and at the University (Nitko, 1998). However, nothing concrete has to date been done to implement such a course.

## 2.11 CONCLUSION

The structure of the formal education is two years for pre-primary, seven years for primary, three years for junior secondary, two years for senior secondary and two-four years of tertiary education. The least developed level is the pre-primary, i.e., the foundation of education, with about 51% of teachers being trained. The current transition rate from junior to senior schools stands at 67%, and Agriculture has the highest number of students among the optional subjects. Agriculture is classified as a Creative, Technical and Vocational subject, and is the only subject in that group which takes the minimum number of learners of 30, while other subjects in the same group take a maximum number of 20.

Agriculture is assessed by three papers. The one that takes the longest time contributes the least (20%) to the final grade, apparently due to the difficulties of ascertaining its validity and reliability. The assessment in this paper is based on two components, namely the practical tests and the project. Assessment of the practicals is ill-structured, while that of the project is well structured. The next chapter discusses how reliability and validity of performance assessment can be improved.

## CHAPTER THREE

### LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

#### 3.1 INTRODUCTION

An extensive literature search was undertaken of primary and secondary sources, including books, paper-based and electronic journals, databases, and conference proceedings and conference papers. The literature review began with an internet search and proceeded through *ERIC* and the Universities of Pretoria and Botswana databases. Key words used in the search were assessment, assessment for learning, authentic assessment, performance assessment, constructivism, pragmatism, validity of assessment, reliability of assessment, formative assessment, and quality assurance in performance assessment.

The literature review was guided by two main research questions, which sought to find out:

- 1. How valid and reliable are the performance assessment processes in Botswana?*
- 2. How can quality assurance processes be developed in order to produce valid and reliable marks for BGCSE Agriculture performance assessment?*

The next Section, 3.2, gives a brief background of the origin of performance assessment, followed by conditions for performance assessment in Section 3.3. Section 3.4 outlines quality assurance of performance assessment internationally. Issues in performance assessment are outlined in Section 3.5 after which conditions of performance assessment in Botswana are delineated in Section 3.6. The discussion focuses on how validity and reliability of performance are ensured internationally in Section 3.7. The conceptual framework of the study is delineated in Section 3.8. The conclusion/synthesis of literature review is presented in Section 3.9.

### 3.2 THE ORIGINS OF PERFORMANCE ASSESSMENT

Performance assessment has been in existence for a long time. Madaus and O’Dwyer (1999) trace the origins of testing to China where it was applied to different disciplines, such as Letters, Law, History, Rituals and Classical Study. It was applied to Education around the early eighteenth century (Morris, cited in Johnson et al., 2009), mainly as oral examinations, and replaced by essay examinations around 1845. Airasian and Russell (2008) posit that “performance assessment has been used extensively in classrooms for as long as there have been classrooms” p. 205.

During those early years, the judgement of the examinees’ performance was mainly qualitative (Hoskin, 1979 Johnson et al., 2009), which introduced the problems of subjectivity and partiality, especially when high-stakes decisions were made. Apart from subjectivity, other problems are such as unreliability of scoring the essay exams were common (Starch & Elliot, cited in Johnson et al., 2009). This did not go unnoticed as the search for better ways of assessing commenced, leading to the invention of multiple-choice tests in 1915 by Frederick Kelly. The invention of multiple-choice testing prefaced the development of standardised, norm-referenced tests (Madaus & O’Dwyer, 1999) and marginalised performance assessment.

During the 1980s and 1990s, the assessment community witnessed the resurgence of performance assessment in education (Johnson et al., 2009). As testimony to this, Stiggins (1995) titled his textbook on performance assessment, *An Old Friend Rediscovered*, in reference to performance assessment. Today performance assessment plays an important role in examinees’ lives, as assessment bodies and commercial examination providers embrace the incorporation of performance assessment marks in certification (Berry, 2008). Clauser, Harik and Margolis (2006) write that performance assessment has increasingly been used as part of high-stakes testing programmes during the past decade, because in some situations it is inevitable.

### 3.3 CONDITIONS FOR PERFORMANCE ASSESSMENT

The distinct characteristics of performance assessment which warrant its implementation are discussed in the subsequent subsections. Performance assessment requires different conditions from those of a paper-and-pencil test. These conditions are such as assessment for learning, assessment enhancing abstract and creative thinking, assessment of authentic tasks, catering for students' cognitive differential development, and complex content which encourages critical thinking.

#### *Assessment for learning*

Research in assessment has traditionally been concerned with studies of the validity and reliability of the externally designed and administered tests and examinations, which were held in high esteem (Black 1993; Harlen, 1994; Popham, 2005). Those for performance assessment were and are still not given much attention. However, the purpose of learning and assessment has since changed from selection, guidance, and prediction of future performance (Stiggins, 2002) to accountability of the school and the education system as a whole (Airasian & Abrams, 2002). This heralded the switch from a testing culture to assessment which focused on encompassing evaluation of learning progress by the learner (Gasemann, 1993), with the provision of useful feedback to learners being the hallmark of assessment for learning (Nitko & Brookhart, 2007; Thorndike & Thorndike-Christ, 2010).

According to Assessment Research Group (ARG) (2002), Assessment for Learning as opposed to Assessment of Learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there. Assessment for learning gives the kind of challenges, diversity and flexibility that make assessment more realistic and educative rather than testing which simply audits learning (Wiggins, 1998):

If we want to improve education to advance our standard of living, we must do away with testing and embrace assessment. Testing is characterized by secrecy and security. When tests are administered, rigid rules are followed (p. 14).

Nitko and Brookhart (2007) posit that educative assessment is authentic and involves showing students by doing, which motivates them to perform better than the awarding of marks. Such assessment has proved to be a powerful school improvement tool, as well as raising students' achievement to unprecedented levels. This led to Neill and Medina (1992) and later Wiggins (1998) to advocate for the abolishment of simultaneous group administration of paper-and-pencil tests which do not take into consideration students' readiness. A decade later, Lissitz & Schafer (2002) supported the reduction of emphasis on large-scale testing. Performance assessment allows students to demonstrate in a variety of ways their understanding, using knowledge and skills learnt from different areas. Diez (2002) concurs with Nitko and Brookhart and Wiggins, but proposes the balancing of classroom-embedded assessment with high-stakes measures.

Research summarised by Black and William (1998) shows that student self-assessment skills, learned and applied as part of formative assessment, enhances student achievement. In 1994, the Averno faculty (McMillan, 2000, p. 211) identified ten elements which when applied in assessment for learning could enhance students' developmental learning processes:

1. Explicit outcomes - clear picture of expectations of candidates' knowledge and performance through the outcomes that guide course and programme development through the dozens of assessment that candidates complete.
2. Performance - assessment of candidates' performance of what they can do with what they know.
3. Public, explicit criteria - criteria which describe the expected quality of performance and must be met.
4. Feedback - feedback from assessors or peers about weaknesses and strengths and how to improve.
5. Self-assessment - assessing oneself, which helps one to become his/her own coach and critic.

6. Multiplicity - assessing more than once using a variety of assessment methods and contexts over time.
7. Externality - trying out in real world situations or bringing others to help assess so as to avoid subjectivity.
8. Developmental nature - assessment which fits the candidates' developmental ability and knowledge.
9. Cumulative nature - assessment which is continuous to give a clear picture of knowledge and skill of students. Students grow over time, therefore are bound to show an improvement.
10. Expansive nature - assessments are developed to elicit from the candidate the most advanced performance of which each is capable.

However, there is no evidence suggesting adoption of these in Botswana as evidence by Mogapi & Yandila (2001) and Yandila, Komane & Moganane (2003) presented in Section 3.6 suggests the contrary.

*Encouraging complex, abstract and creative thinking*

There is no single correct answer to real-world problems, and standardised testing follows rigid regulations which could lead to failure by students to engage in demanding creative tasks (ARG, 2006; Shepherd 2000, 2008). In most cases, standardised paper-and-pencil tests are of low-order measuring how much learning has taken place with little regard to context, creativity and processes. Assessment as a social construction should consider students' social background and students' prior knowledge. For example, a *pest* in one region could be an extremely valuable creature in another region. Setting a multiple-choice question on such a "pest" could disadvantage students from a region where it is not regarded as a pest, so a question on *pest* should rather be set to encourage creative and thoughtful application and meaningful use of knowledge to solve problems caused by pests.

Assessment in performance tasks provides the opportunity to assess thinking processes that the students undergo to construct their responses (Airasian, 2005). The assessor observes the students performing a task to find an answer to a problem. The teacher then marks the students' every step taken and guides the student to the right procedure whenever the student deviates. The intention is not to assess the ultimate answer, but how the answer was arrived at (Airasian & Russell, 2008; McMillan, 2004). The former is the goal of selection type of assessment. It assumes that when the student gets an item correct, the student must have followed the correct process, but there is no direct evidence to support the assumptions. As such, performance assessment helps to gauge what pupils can do as opposed to selection tests that assess what pupils know (Neill, & Medina, 1992).

#### *Authentic Assessment*

Performance assessment should resemble the activities taking place in the real world (McMillan, 2000). According to Diez (2002), Rennert-Ariev (2005) and Ryan (2006), performance tasks address demanding tasks which normally span longer periods, hence requiring students to use many different skills and abilities. For example, students growing *a crop* spend at least four months managing it, and during that period they are engaged in a number of management activities. In carrying out these activities students are required to apply knowledge and skills acquired from different areas, including affective skills.

This kind of learning is authentic in nature (Johnson et al., 2009; Nitko & Brookhart, 2007) because students perform in the context of the real-world situations in which the skills are to be applied. They are involved in doing rather than just knowing how to do it or simply know it. Authentic skills are not fixed, hence they cannot be assumed to be conducted under standardised conditions or manifest themselves always in the same way at any time across contexts (Airasian, 2005; Popham, 2005). This calls for various formats and methods to be used for assessing students (Stiggins, 1997).

Wiggins (1998) developed a set of six standards for judging the degree of authenticity in assessment:



- a) *Is realistic*: the task replicates the ways in which a person's knowledge and abilities are tested in real world situations.
- b) *Requires judgement and innovations*: the student has to use knowledge and skills wisely and effectively to solve unstructured problems, and the solution involves more than following a set of routine procedures or plugging in of knowledge.
- c) *Asking the student to do*: the student has to carry out the exploration and work within the discipline of the subject area, rather than restating what is already known or what was taught.
- d) *Replicates or simulates the context in which adults are tested in the workplace, in civic life, and in personal life*: contexts involve specific situations that have particular constraints, purposes, and audiences. Students need to experience what it is like to do tasks in the workplace and other real life contexts.
- e) *Assesses the student's ability to efficiently and effectively use a repertoire of knowledge and skills to negotiate a complex task*: students should be required to integrate all knowledge and skills needed, rather than to demonstrate competence of isolated knowledge and skills.
- f) *Allows appropriate opportunities to rehearse, practice, consult resources, and get feedback on and refine performances and products*: rather than relying on secure tests as an audit of performance, learning should be focused through cycles of performance – feedback – revision - performance, on the production of known high-quality products and standards, and learning in context (p. 22-24).

However, in some situations, conducting authentic performance assessment is unattainable, such as when performance is complicated and equipment is expensive, or puts other people's lives in jeopardy. For example, the application of chemicals to control pests by students under the age of sixteen is not legally allowed. In such situations, simulations could be an alternative (McMillan, 2004) to serve as an intermediate step to performance that involves a higher degree of realism.

### *Catering for different developmental rates*

Students have been found to develop intellectually at different rates, depending on their background, experiences and learning styles (Neill, & Medina, 1992). Since learning is related to intellectual development, it should follow therefore that learning should also be differentiated to cater for individual differences. The multidimensionality of students' development requires learning to be based on theories that encompass dimensions of cognitive, psychomotor and affective skills (Nitko & Brookhart, 2007). Such learning provides an opportunity to students who do poorly on the cognitive dimension to show their achievement in performance assessment (Airasian, 2005). Assessing students in multiple ways provides the opportunity for students to engage in performance assessment, which renders equal opportunity to be assessed in all domains of development. Research has found that individuals exhibit different ways of knowledge and problem solving that reflects different styles, not different abilities, yet standardized paper-and-pencil tests assume that all individuals perceive information and solve problems in the same style (Neill & Medina, 1992).

### *Covering complex content*

Performance assessment covers in-depth content of knowledge and skills (Johnson et al., 2009), the coverage of which comes with the problem of relatively few tasks being used as compared to other formats of assessment, resulting in scores suffering from external validity (Lane & Stone, 2006). Airasian (2005) advises that this could be overcome by properly created performance assessment tasks which sample a wide range of abilities to be applied by the student in solving complex problems. Wiggins (1998) cautioned against the use of low-order thinking skills as the solution to external validity. Performance assessment allows for the assessment of students' complex processes, as well as their product. Assessment of processes is crucial for the accomplishment of quality products and activities that have momentary evidence that cannot be formatively assessed by paper-based tests (Black, 1995; William & Black, 1996). These thought-provoking tasks have multiple solutions to allow students to construct their own meaning fostering development of thinking in varied styles.

In conclusion, performance assessment tasks should be essential, drawing from the core curriculum and representing a “bigger idea”. The tasks should be authentic, using processes appropriate to the discipline, and students should value the outcome of the tasks which in turn lead to problems that require them to draw from deeper faculties in solving, rather than replicating known procedures. Quality assurance processes should be in place to validate performance assessment marks. Having looked at the conditions of performance assessment, the discussion now focuses on international examples of quality assurance.

### 3.4 QUALITY ASSURANCE OF PERFORMANCE ASSESSMENT INTERNATIONALLY

Performance assessment has become a necessary undertaking for many examination boards, with the main focus on quality assurance (Khoo & Idrus, 2004; Maughan, 2004), defined by Oakland (1993, p. 13) as “broadly the preventing of quality problems through planned and systematic activities (including documentation)”. Performance assessment is premised on entrenching quality in the system and continual auditing and reviewing (Walklin, 1992; Doherty, 1994).

Although school-based performance assessment has been criticised for its lack of reliability in particular (Chong, 2009), it is necessary to maintain the right balance between teachers’ professional judgment and national testing for national assessment systems to be comprehensive, rigorous and meaningful, while at the same time improving teaching and learning (Queensland Studies Authority, 2009; Pellegrino, Chudowsky & Glaser, 2001). All the same, there is no single solution in achieving this, and different countries employ different strategies or the same strategies applied differently, as per the dictate of their contextual factors (Broadfoot, 1994; Maxwell, 2004; Raivoce & Pongi, 2000). Despite the combinations and permutations possible, the first step in ensuring quality in performance assessment is to embed quality into the processes (Campbell & Rosznyai, 2002; Richard, 1993), which includes but not limited to teacher development of tasks, training to assess; resources provision; leadership commitment; development of

learner/support materials, moderation, authentication, internal monitoring, external monitoring and supervision, multi-rating, and school approval, (Chong, 2009; Khoo & Idrus, 2004).

Training of teachers to acquire the appropriate expertise is essential (Broadfoot, 1994), as the public has confidence in trained teachers to conduct assessment professionally and ethically (Maxwell, 2004). Germany and Australia are well known for emphasising professional development of teachers to assess (Broadfoot, 1994; Queensland Studies Authority, 1998), because their assessment procedures are largely the responsibility of teachers, even for certification and selection purposes, with minimal external intervention or moderation (Gasemann, 1993).

Teachers in Germany and Australia develop their own good quality assessment tasks and procedures, therefore assessment marks are not aggregated by a mathematical formula to produce an overall result. Rather, the result involves an interpretation of the final product of the student's work by a judgement of the standard it demonstrates when compared to a set of grade descriptors (Mercurio, 2008; Maxwell, 2004). Teacher training is not a sufficient condition but high management commitment to quality provision can enhance their performance (Burdett & Johnson, 2009). The management should be well-versed in assessment practices to manage assessment process and support teachers (Bennett & Taylor, 2004; Calvo-Mora, Antonio Leal & Roldan, 2006; Wild & Ramaswamy, 2008).

The importance of moderation in ensuring comparable outcomes and improving teachers' assessment capabilities through applying agreed standards consistently by the individuals involved is very important (Queensland Studies Authority, 2009). According to Klenowski and Wyatt-Smith (2008): "Moderation can no longer be considered an optional extra and requires system-level support especially if, as intended, the standards are linked to system-wide efforts to improve student learning." (p. 1). Jordan and McDonald (2008), Masters and McBryde (1994), and Stanley and Tognolini (2008) note that the inter-rater reliability of the moderation system practiced in Queensland, in which teachers and schools were accountable for the assessment and reporting of student achievement, surpassed that of many external examination regimes. Furthermore, Bennett

and Taylor (2004) assert that a system of moderation of teachers' judgments through professional collaboration benefited teaching and learning as well as assessment. Such a moderation procedure has more than a quality assurance function.

Moderation as a social ratification of teachers' assessment (Radnor & Shaw, 1995) is directed towards ensuring that quality assessment standards have been applied consistently. Moderation directed at ensuring quality is the most commonly applied form of moderation in Britain, New Zealand, Malta, Kenya, and Australia, to mention only a few countries (Boustead, 2008; Broadfoot, 1994; Harlen, 1994; Maxwell, 2004; Onyango & Ndege, 2007; Raivoce & Pongi, 2000). However, moderation employing a variety of methods that combine both quality assurance and quality control procedures has been found to yield better results (Berry, 2008; Keightley & Coleman, 2002; Queensland Studies Authority, 2009; Maxwell, 2004; Raffan, 2000; Raivoce & Pongi, 2000). For example, members of the markers' panels of different subjects visit the schools to moderate the candidates' coursework (Grima & Ventura, 2000; Queensland Studies Authority, 2008).

In SPBEA, a two-level moderation procedure is conducted, using students' samples to account for any differences between schools within a country and between countries, while statistical moderation is carried out on Teacher Designed Tasks. In the USA, the assessment is validated and calibrated using three models, namely: i) a national exam for a sample of students in each grade level, used to verify standards assessed by regional or local examinations; ii) some element of national exam with local exam; or iii) marking visiting teams cross-moderate between schools (Broadfoot, 1994).

While a few countries still moderately employ statistical moderation, such as Sweden, Hong Kong, South Africa, and to some extent New Zealand, (Berry, 2008; Broadfoot, 1994; Lennox, 2000; Singh, 2004), majority have abandoned its use (Broadfoot, 1994; Harlen, 1994; Maxwell, 2004; Radnor and Shaw, 1995; Raivoce & Pongi, 2000), in favour of embracing moderation processes that ensure quality (Boustead, 2008; Keightley & Coleman, 2002; Lennox, 2000), Critics of statistical moderation argue that this constitutes a typical misuse of statistical tools:

the choice of a theory examination as reference standard for the moderation of practical grades is not beyond criticism. This inadequate and lazy practice is no longer allowed and statistics are now used in support of more relevant techniques of moderation, (Kempa, 1986, p. 85).

statistical moderation, which often uses as its external reference point a written public examination, can stifle innovation in the classroom, and, in particular, can whittle away the professional skills of the teacher to design the assessment and make appropriate judgments (Mercurio, 2008, p. 9).

School approval or accreditation is another way of ensuring quality in performance assessment (Council for Higher Education Accreditation [CHEA], 2002), as it involves an evaluation of the capacity of a school to enter students for the board's qualifications, and provides an opportunity for schools to improve their management, learning and assessment processes. Before the school is allowed to conduct performance assessment, an audit of its capabilities to successfully implement performance assessment is conducted (CHEA, 2002; Colbeck, Caffrey, Donald, Lattuca, Reason, Strauss, Terenzini, Volkwein, & Reindl, 2000; Jones, 2002). Officers from the examination board visit schools during the year in order to evaluate the physical and human resources available to carry out the coursework as specified in the syllabus, and the type and standard of the coursework.

Officers also observe and evaluate the assessment methods and procedures (Grima & Ventura, 2000). Participating schools are required to submit an assessment programme, clearly indicating what they intend to do, as well as the various assessment tasks that make up the programme for each subject that has a performance assessment component (Keightley & Coleman, 2002; Raivoce & Pongi, 2000). For example, SPBEA has to check for compliance with such factors as prescribed requirements, appropriate standards, and timeframe (Raivoce & Pongi, 2000) before any programme is approved for implementation.

After the school has been given permission to implement performance assessment, monitoring is carried out regularly to ensure that assessment is carried out in a

satisfactory manner and that the school complies with the specified standards. Schools are required to maintain quality assurance systems to continue to carry out the Board's qualifications, and monitoring of adherence to standards is done both internally and externally. Selected schools are visited each year to verify that internal assessment programmes are being followed and to assist teachers in the delivery of the learning programmes (Keightley & Coleman, 2002).

One fundamental aspect of ensuring quality in performance assessment is by developing task frames to guide the development of tasks (Keightley & Coleman, 2002). Task frames are summary prose statements which detail the types and range of performance, representing different levels stipulated in the national curriculum. In Germany, France and Australia, performance tasks are developed by teachers themselves, after undergoing vigorous training (Broadfoot, 1994; Keightley, 2002), whereas in Sweden and Singapore task development is highly centralised (Chong 2009; Maughan, 2004).

The SPBEA, for example, uses three approaches to developing tasks: i) centrally developed tasks known as Common Assessment Tasks; ii) Teacher Designed Tasks (TDTs) developed by individual teachers in school; and iii) Common Assessment Frame tasks (CAFs), all determined by SPBEA but the tasks of which are developed by teachers (Ravoice & Pongi, 2000) then evaluated according to the achievement standards pre-defined to judge students' ability to meet the expected level of assessment (ARG, 2006; Keightley & Coleman, 2002). The details of developing tasks shall be discussed in detail in Section 6.3.

Scoring of performance assessment presents a problem as it sometimes involves judgement. The subjectivity of performance assessment is greatly reduced if multiple rating is used (Airasian, 2005; Airasian & Russell, 2008; Thorndike & Thorndike-Christ, 2010), with research having proved that the same test could be scored differently by different teachers and that even the same teacher could score responses differently at different times (Rennert-Ariev, 2005). According to Rudner and Boston (1994), multiple ratings improve reliability, in as much as multiple test items can improve the reliability of standardised tests. Further improvement of the reliability can be made by the use of

criteria in scoring (Nitko, 2004). Throughout the process of scoring, scorers recalibration of raters through refresher practice sessions (Johnson et al., 2009), to avoid ‘rater drift’ should be done (Becker & Pomplun, 2006).

It is evident from the above discussion that the majority of countries were moving towards embracing performance assessment, to compliment standardised testing. While in some countries performance assessment tasks were centrally developed, they were solely the responsibility of the teacher in other countries. Similarly, the magnitude of quality insurance fell along the continuum of professional development of teachers to moderation. Forms of moderation applied ranged from the controversial statistical moderation to the consultative visiting moderation (Radnor & Shaw, 1995). In other countries, the implementation of performance assessment is fully developed to an extent of completely depending on the assessment conducted by teachers for certification, without any moderation or external intervention.

Though performance assessment allows the teacher to provide the information about what the student can do, its conduct remains problematic. The general public has not yet accepted it as a formal way of impartially assessing students, and even professional teachers seem not to understand their role adequately (Chong, 2009). This is because teachers’ pedagogical training does not normally emphasise performance assessment.

Given the above factors that can enhance quality performance assessment the following section discusses issues in performance assessment that hinders its effective conduct.

### 3.5 ISSUES IN PERFORMANCE ASSESSMENT

The past two decades have witnessed a global trend towards performance assessment (Airasian & Russell, 2008; Abraham, 2008; Berry, 2008; Crooks, 2004; Harlen, 1994; Harlen, 2006; Maughan, 2004; Maxwell, 2004; Pongi, 2004; Raffan, 2000; Raivoce & Pongi, 2000). For example, over 40 states of the USA had adopted some form of performance assessment by 2000, (Patchen, 2004), while in the United Kingdom (James 1994) every school curriculum subject has introduced performance assessment as a



component in the past 20 years (Raffan, 2000; Berry, 2008). In the Hong Kong education system, performance assessment was introduced as an important aspect of the assessment reforms (Hamp-Lyons, 2009), while in SPBEA it was introduced as a way of striving to provide quality and timely service to its clients (Pongi, 2004). In India, the 2005 National Curriculum Framework proposed a shift from traditional assessment based on behaviourism to constructivist approaches (Kapur, 2008).

However, performance assessment is not without problems (Chong, 2009). A discussion of the issues that arise in the use of performance assessment follows, looking at the problems, the debates, and how specific countries have dealt with these, as well as issues that have not yet been resolved. Major problems in performance assessment will now be examined in turn.

#### *Development of tasks*

Variation in the demand of tasks or opportunity provided by the tasks undertaken by students is one issue that is problematic in performance assessment (Department of Education 2001, cited in Singh, 2004). The inability by teachers to develop appropriate materials for assessment purposes, consistent with the relevant national curriculum (Kanjee & Sayed, 2008), is due to lack of training (Chong, 2009; Maxwell, 2004; Nenty, Odili and Munene-Kabanya, 2008; Stiggins, 2000). Developed countries are making progress towards entrenching quality in performance assessment among teachers (Broadfoot, 1994; Maxwell, 2004). This is because teachers' technical competence to assess invariably facilitates the interpretation of performance criteria. To prevent the intrusion of irrelevant contextual information in making judgements, marking schemes are to be understood and applied in the same way.

Maxwell (2004) asserts that if teachers are properly trained and given enough support resources, they can design and develop sound assessments, which can then be used to:

- i) determine what a student has learnt and what s/he still needs to learn,
- ii) help each student learn and use knowledge well,

- iii) determine how well the teacher applied an instructional process, and
- iv) provide information to students, teachers, and parents (Mamary, 2007, p. 188).

#### *Provision of resources*

Implementing performance assessment on a large scale requires massive resources, which are costly (Tindal & Haladyna, 2002), but consequential gains to the learner are immeasurable. Doty (1996) suggests that costs of implementing performance assessment can be significantly reduced by identifying and controlling expenses through budgeting, measuring, and analysis, to achieve higher quality education at lower cost. Since performance assessments are perceived to be expensive, as in the case of portfolio which is developed over a period of a year with many students in a class (Mills, 1996; Johnson et al., 2009; Nitko & Russell, 2007), limited resources and time are often directed towards less expensive standardised testing (Stiggins, 1997). Pellegrino, Chudowsky & Glaser (2001) called for the balance of mandates and resources to be shifted from an emphasis on external forms of assessment to an increased emphasis on classroom formative assessment, given that well-resourced schools tend to perform better (Howie & Plomp, 2001).

#### *Teacher Workload*

Performance assessment as a student-centred approach requires more time for individualised instruction (Fung et al., 1998) and recording of the student achievement and progress. As a result, some teachers view school-based assessment as an extra workload imposed by an external institution (Keightley & Coleman, 2002; Torrance, 1995), which should be paid for particularly when done for summative purposes (Grima & Ventura, 2000). Because of the work involved, teachers prefer externally set practical examinations to school-based assessment (Raffan, 2000), despite the well-documented validity evidence of the latter. Teachers who then engage in school-based performance assessment resort to inflating students' marks under the pretext of time constraint (Raivoce & Pongi, (2000).

It is generally accepted that class size and workload are related, however less clear is whether class size has any effect on achievement. Mixed and inconclusive findings have been reported about the effect of class size, for example Finn and Achilles (1990), Hoxby (2000), Milesi and Gamoran (2006), Nye, Hedges and Konstantopoulos (2002), Pong and Pallas (2001), found little or no gain in small class sizes. On the other hand, Angrist and Lavy (1999), Knostantopolous (2008), Knostantopolous and Chung (2009) contend that small class sizes yield positive results, particularly in developed countries. Finn et al. (2003), Jones (2006), Miller, Sen and Malley (2007) identified gains for small class sizes to be : (i) more participation, engagement and identification; (ii) more teacher time per student; and (iii) more time for individualised assessment and increased time on task.

However, class sizes were found to be large in African and Asian schools (Bery, 2008). For example, in Malawi, class size in 1994 at primary level was 100 (Nowa-Phiri, 2000), while in South Africa grade 8 average Mathematics class size was 46 students (Howie & Plomp, 2001). Large class size was found to be an impediment to implementing authentic assessment (Howie, 2006). Howie and Plomp (2003) found that class size and work load affected students' performance in mathematics. If there are too many students in the classroom, the teacher's assessment focus tends to be on class, or perhaps the small group, rather than the individual student.

#### *Low weightage*

The contribution of performance assessment towards final grade varies significantly from country to country, depending on the development of the structures in place as well as the confidence the public has in performance assessment outcomes (Chong, 2009). Nitko (1995) proposed three models for combining performance assessment results with National examination results:

Model one: using performance assessment only at school level but not counting them toward certification

Model two: count performance assessment toward certification or selection using a compensatory model (e.g. regression weighting)

Model three: count performance assessment toward certification or selection but fix the percentage weight (e.g. 40% or 60% of the total performance assessment: (a) count only the last few years, (b) count all years, or (c) count all years but weigh earlier years less than later years. (p. 5)

A number of countries have reported varying contributions of performance assessment, even between subjects within a country. In England and Wales, for example, significant elements of teacher assessed coursework and practical work was weighted between 20% and 100 %, depending on the subject and syllabus followed (Torrance, 1995). The weight of performance assessment component for PSSC countries ranged from 40% to 100%. (Raivoce & Pongi, 2000; Ventura & Murphy, 1998), while in Germany, performance assessment as the responsibility of teachers contributed 100%, with minimal external intervention (Broadfoot, 1994; Gasemann 1993).

In the UK, performance assessment in the non-core subjects contributes 100% of the final statutory assessment (end of Key Stages 3 and Stage 14), while at the GCSE level varies. For example, Biology's coursework contributes 20% (Maughan, 2004). In Australia, performance assessment's contribution varies from one province to another, between 50% and 100% (Keightley & Coleman, 2002). For example, in Queensland it contributes 100%, hence there has been no standardised public examinations for over 35 years (Maxwell, 2004).

In Kenya, CA was implemented in only three subjects and ranged from 10% to 25% (Noor, 2008), while in Namibia, summative CA contributed between 30% and 50% to the end-of-course grade (van der Merwe, 2000). Njabili (1987) reported CA's contribution to be 50% across the board in Tanzania. In South Africa, CA contributed 25% to the final Grade 12 examination grade (Kanjee & Sayed, 2008; Singh, 2004; Van der Berg & Shepherd, 2010).

### *Teacher training*

It has been discussed under 'task development' that assessment should be all-encompassing. Research has revealed that teachers lack skills to develop tasks that can

recognise full range of achievements of all students, despite the fact that they are the appropriate assessors of what is inaccessible to the external examination (Pellegrino, Chudowsky & Glaser, 2001; Tindal & Haladyna, 2002; Wiggins, 1998). Nevertheless, building the capacity and competency of teachers to carry out assessment in the classroom effectively and consistently is a challenging task (Chong, 2009). Due to inadequate training, teachers are not prepared to assess their pupils, especially on performance tasks (Kellaghan & Greaney, 2003).

Howie (2006) pointed out that one of the reasons teachers in South Africa cannot implement performance assessment successfully is because they have had insufficient training in assessment. Stiggins (2002) noted that about only one quarter of states in America require that pre-service teachers take an assessment course, but only three states require competence in assessment as a requirement or condition of being licensed as a principal, while no state certifies that competence.

Lack of training in assessment results in teachers deemphasising or neglecting untested materials (Tindal & Haladyna, 2002), as well as testing students on trivial outcomes which seek to find out whether the child knows, understands or can perform predetermined tasks (Torrance & Pryor, 1998). Wiggins (1998) is of the view that students need to be given the same training that the assessors receive, so as to be able to judge whether their work is up to standard. Despite lack of understanding of principles of assessment by teachers, they spend more than half of their professional time involved in assessment-related activities (Stiggins, 1997; Boyle & Christie, 2000). This prompted Pellegrino, Chudowsky and Glaser (2001) to declare that instruction on how students learn and how learning can be assessed should be a major component of teacher pre-service and professional development programmes.

#### *Teacher role conflict*

Teacher training in assessment is important, but on its own is inadequate since the teacher is required to play a dual role of facilitator and assessor of his/her students (Keightley, 2002; Keightley & Coleman, 2002). Chong (2009) points out that such a situation subjects the teacher to a serious challenge because s/he cannot suppress one when

engaged in the other. As a consequence, one of the most frequent concerns about school-based assessment is the issue of teacher bias (Keightley & Coleman, 2002), and it is not surprising to find low variance and a skew towards high marks (Grima & Ventura, 2000).

Some argue that high marks are expected since students are guided by their teachers during the learning process, and are encouraged to improve their performance before they are awarded the final mark for their work (Maxwell, 2004). The portfolio assessment is the closest example in which initial tasks are given low weightage comparatively (Nitko & Brookhart, 2007), or a few best experiments being chosen and scored, as is the case in Malta (Grima & Ventura, 2000). All these tend to affect some teachers' judgements. In some instances, teachers are affected by physical attractiveness of students, by aspects of behaviour or perceptions of ability (Raffan, 2000), and so award or deny marks where they are due.

#### *Lack of confidence in internal assessment*

Stiggins (1995) reported resistance towards introduction of teachers' classroom assessment. There is a widely held perception that any examination where external examination does not feature strongly is unreliable and biased (Broadfoot, 1994; Chong, 2009; Keightley & Coleman, 2002; Stiggins, 1997), and even in countries such as Australia, where performance assessment has been in existence since 1970, this perception is still entrenched (Keightley & Coleman, 2002). Of late, there is a paradigm shift towards embracing performance assessment, although many are still equating assessment to external examinations (Raivoce & Pongi, 2000). Lack of confidence by the public is borne from the public's lack of understanding of basic principles of appropriate test interpretation and use (Pellegrino, Chudowsky & Glaser, 2001).

#### *Plagiarism*

Plagiarism is one of the challenges in performance assessment (Pongi, 2004), the most common forms being:

1. word-for-word copying of sentences or paragraphs from one or more sources which are the work or data of other persons (including books, articles, working

- papers, conference papers, websites or other students' assignments), without clearly identifying their origin through appropriate referencing.
2. closely paraphrasing sentences or paragraphs from one or more sources without appropriate acknowledgment in the form of a reference to the original work or works.
  3. submitting work which has been produced by someone else on the student's behalf as if it were the work of the student.
  4. producing work in conjunction with other people (other students, a tutor, parents) when it is purported to be work from the student's own independent research. (<http://www.griffith.edu.au/>).

The list is not exhaustive, as Maxwell (2004) argues that even work that is refined and resubmitted on the basis of teacher feedback may constitute plagiarism, since it is difficult to separate the student input from that of the teacher. In other examination boards, validation is through authenticating the marks on the final form (Grima & Ventura, 1998), and when in doubt the candidates are called for an interview to establish if the work was copied or recycled. However, it is not always possible for teachers to realize that the work presented is not original.

All these problems consequently lead to low validity and reliability of performance assessment, which is discussed in Section 3. 7.

### 3.6 THE CONDUCT OF PERFORMANCE ASSESSMENT IN BOTSWANA

The recommendation to incorporate performance assessment in the final grade was made in the First Commission on Education of 1977 (Government of Botswana, 1977), and reiterated by the Second Commission in 1993, resulting in the Revised National Policy on Education (RNPE) of 1994. Following the second recommendation, the Examining Board formed a task force in September 1993, comprising Ministry of Education and the Examination Body Officials, to consolidate needs assessment for basic education in

Botswana. In 1998, the Examining Board engaged a consultant to report on the logistics and modalities of implementing performance assessment.

Both the Task Force and the consultant recommended the introduction of Criterion-Referenced Testing (CRT) with diagnostic capability; development of Continuous Assessment (CA) procedures for all school grades to be used as part of final examinations results; and development of materials and training programmes in CRT and CA during pre-service and in-service teacher training, as well as training Ministry of Education personnel such as Principal Education Officers (PEO) (Nitko, 1998). It can be reported that no further work has been done since, and there is no policy on performance assessment other than subject-specific procedures.

Currently, performance assessment is limited to practical subjects and quality is assured through visiting moderation at the end of the coursework, where the teacher and the moderator reconcile their differences (Radnor & Shaw, 1995). Statistical moderation is applied in Design and Technology, in addition to visiting moderation. External moderation is preceded by internal moderation in case more than one teacher was involved in marking.

The issue of payment for performance assessment as experienced in other parts of the world (Grima & Ventura, 2000) is not exceptional to Botswana. Teachers argue that performance assessment increases their already high workload. The issue is so serious that, in 2008, Teachers' Unions took the government to court, demanding to be paid for conducting performance assessment used for external purposes or it is removed from their mandate. The court ruled in their favour. As of September 2009, Teacher Unions have instructed teachers not to submit performance assessment marks until the Examining Body has agreed to pay (*Mmegi* Newspaper, 26 October, 2009, p. 4). This development negatively affects the public's confidence in teacher assessment for certification.

The weight of performance assessment is very little, ranging from 20% in Science subjects and Agriculture (MoE&SD, 2009), to 50% in the majority of subjects. Only Art and Design is assessed 100% by performance assessment (MoE&SD, 2001). The low



weightage compares well to other African countries, such as Kenya, Namibia and Tanzania (Njabili, 1987; Noor, 2008; van der Merwe, 2000). Thobega and Masole (2008) attributed the low contribution by Agriculture performance assessment to its questionable reliability. On the other hand, a study by Rathedi (1987) pointed to the need to increase performance assessment contribution towards the final grade. For example, lecturers (78.7%) and graduates (96.7%) did not embrace the testing mode to an excessive degree at the expense of contextualized learning going on throughout the course of study. Rathedi did not outline how quality would be assured for performance assessment to be valid and reliable.

External monitoring and supervision regarding performance assessment is not sufficient, as a result of confusion as to whose responsibility it is among the four departments<sup>7</sup> of the Ministry of Education in Botswana. Mogapi and Yandila (2001) and Yandila, Komane and Moganane (2003) summarised the problems of teaching and conducting performance assessment in Botswana to be: large class sizes of up to 40 students; large teaching loads; absence of laboratory assistants; lack of exemplary teaching materials; inadequate training to carry out coursework assessment; and insufficiency of teachers' orientation on appropriate teaching methods.

Based on the findings of Sections 3.5 and 3.6, quality assurance processes for performance assessment between Botswana's and International practice are summarised in Table 3.1.

Table 3.1: *Comparison between Botswana and international practice on quality assurance processes for performance assessment*

<b>Characteristic</b>	<b>International practice</b>	<b>Botswana practice</b>
Teacher training	Advanced to the extent that teachers develop their own good quality assessment tasks and procedures,	Not emphasised
Moderation	Result involves an interpretation of the final product of the student's work by a judgement of the standard it demonstrates when compared to a set of grade descriptors. Moderation directed at ensuring quality by using a variety of methods that combine both quality assurance and quality control procedures.	One moderator with an outsider perspective. Moderation is a one-off activity by one person directed at controlling quality at the end of the process.
Accreditation	School are accredited and visited during the year to evaluate the physical and human resources, assessment methods and procedures. Participating schools are required to submit an assessment programme.	Schools are inspected once at the beginning when it applies to offer the subject.
Monitoring & Supervision	Monitoring of adherence to standards is done both internally and externally by visiting schools each year and to assist teachers in the delivery of the learning programmes.	Monitoring internally is not rigorous. External monitoring is not common.
Workload	Small class sizes.	Large class sizes
Development of tasks	Tasks are developed either by teachers after undergoing vigorous training or centrally developed. Tasks are of high quality.	No task frames or centrally developed tasks. Every school develops its own task.
Scoring	It is done by multiple raters.	It is done by one rater
Assessment Instrument	Use of detailed clearly written criteria.	Individual schools or even teachers develop their own
Weight	High	Low

### 3.7 VALIDITY AND RELIABILITY OF PERFORMANCE ASSESSMENT INTERNATIONALLY

The issue of validity and reliability in performance assessment is topical (Burger & Burger, 1994; Chong, 2009; Cizek, 1991; Kane, 2008; Mehrens, 1992; Messick, 1989). While validity and reliability of standardized norm referenced testing is well established (Stobart, 2008), that of performance assessment is not (Hargreaves, 2007). For performance assessment to provide credible outcomes there should be no compromise on their validity and reliability (Linn et al., 1993; Mehrens, 1992). Given that absolute validity and reliability are almost impossible to achieve even in written examinations (Harlen, 1994), van der Merwe (2000) implores psychometricians to adopt a lenient stance toward accepting lower levels of validity and reliability.

Comparatively, performance assessment has been found to rate highly for all aspects of validity (Linn et al., 1991), but there have often been significant problems with reliability (Broadfoot, 1994). The claim of established validity of performance assessment was countered by Cizek (1991) and Mehrens (1992), who argued that this only applies to face validity, and so pertains only to what the test appears superficially to measure. Since reliability can be more readily evaluated and quantified than validity, reliability is persistently emphasised, even at the expense of validity (Raffan, 2000). However, Woods (1991) and William (1992) adopted a compromise stance of a trade-off between reliability and validity for any national system of examinations.

#### 3.7.1 *Validity*

Nichols and Williams (2009) purport that the concept of validity has evolved over time from the validity of an instrument (Ary, Jacobs, Razavieh & Sorensen, 2006) to the interpretation, meaning and usefulness of the scores derived from the instrument (Ary et al., 2006; Salvia & Ysseldyke, 1998; Yao, Thomas, Nickens, Downing, Burkett & Lamson, 2008). This evolution was emphasised by Ary, et al (2006) when they wrote: “validity does not travel with the instrument” (p. 243).

Although Lissitz and Samuelsen (2007) are still of the view that validity is the property of the test, independent of any proposed interpretation or use of the results, most textbooks even today talk of the validity of the instrument. This certainly does not apply to validity in qualitative studies, where it is addressed through honesty, depth, richness and scope of data achieved (Mertens, 2010). There are several different kinds of validity, but only a few applicable to this study are outlined, namely internal validity, external validity, content validity, construct validity, criterion-related validity and consequential validity (Cohen, Manion & Morrison, 2000).

#### *Internal validity*

It is the intention of any study to maintain a high degree of internal validity, that is, the observed changes in the dependent variable which are due to the effect of the independent variable, and not to some other extraneous or lurking variables (Mertens, 2010). Aiken (1996) asserts that internal validity is akin to reliability (p.65), and in qualitative research it is assured through credibility, dependability, conformability, and authenticity of data (Mertens, 2010). To improve internal validity of the study, threats have to be eliminated (McDavid & Hawthorn, 2006), through strategies suggested by Cohen, Manion and Morrison (2000) which include: triangulation of data collection methods; using participant researchers; using mechanical means to record, store and retrieve data; using peer examination of data, and persistent observation. Furthermore, the authors argued that threats to internal validity in qualitative research are built in, since it is assumed that they will happen.

#### *External validity*

External validity refers to the degree to which results can be generalised to the wider population, cases or situations (Aiken, 1996; Cohen, Manion & Morrison, 2000) based on the assumption that the sample is representative of the population (Mertens, 2010). To achieve external validity in quantitative research, variables have to be controlled, and samples randomized, whilst for qualitative research human behaviour is infinitely complex, irreducible, socially constructed and unique (Cohen, Manion & Morrison,

2000). Yin (2009) suggests that the use of multiple cases can strengthen the external validity of results.

External validity in qualitative research is interpreted as comparability and transferability (Guba & Lincoln, 1989; Lincoln & Guba, 1985), thus data in qualitative research can be translated into different settings and cultures. If clear, detailed and in-depth description of research is made then others can decide the extent to which findings from one piece of research are generalisable to other situations. Lincoln and Guba (1985) and Bogdan and Biklen (1992) caution researchers that it is not their task to provide an index of transferability, but rather to provide thick description (Mertens, 2010) of the settings, people, and situations to which they might be generalised.

#### *Content validity*

Whether data is collected using either adopted or adapted instruments or developing one's own, determination of content validity is an important first step (Viswanathan, 2005). In judging content validity, the content domain, which includes both the subject matter and the type of behaviour or task desired from students (Mehrens & Lehman, 1991; Aiken, 1996; Moskal, & Leydens, 2000; McIntire & Miller, 2007), and universe of situations must first be defined, and thorough inspection of the items made. Recently, Kane (2008) redefined content validity to include both judgments about content and some analysis of reliability and scaling issues. It should be emphasized that an instrument may have high content validity for one user and low content validity for another, because they wish to infer to different domains (McIntire & Miller, 2007).

#### *Construct validity*

Thorndike and Thorndike-Christ (2010,p.11) define a construct as an abstract. Construct is therefore the most important and most difficult form of validity to establish (McIntire & Miller, 2007; Viswanathan, 2005), hence the construct should be operationalised. Fink (2005) simply defines construct validity as a measure that distinguishes between people who have certain characteristics and those who do not. An instrument is said to have

construct validity if the instrument results are in keeping with this expectation (Devitt, Kurrek, Cohen, & Cleave-Hogg, 2001).

There are two strategies for demonstrating construct validity, namely convergent and discriminant (McDavid & Hawthorn, 2006; McIntire & Miller, 2007; Viswanathan, 2005). In convergent validity constructs that should be theoretically related are indeed related, while in discriminant validity constructs that are not supposed to be linked are not correlated (McDavid & Hawthorn, 2006; McIntire & Miller, 2007). As with internal validity, construct validity is also vulnerable to threats, two of which were identified by Messick (1995) in Mehrens (1991) as being major ones: (1) construct underrepresentation, in which the assessment is too narrow and fails to include important dimensions or facets of the construct; and (2) construct irrelevant variance in which the assessment is too broad and contains excess variance because of intrusion of other constructs.

#### *Criterion-related validity*

Criterion-related validity is a validation method used to determine whether a test indeed predicts what it claims to predict (McIntire & Miller, 2007; Mehrens & Lehman, 1991). A test has evidence of criterion-related validity when it demonstrates that its scores are systematically related to a relevant criterion. Predictive and concurrent validity are two types of criterion-related validity (Ary et al., 2006; McIntire & Miller, 2007). Predictive method is used to forecast future performance (Fink, 2005) while concurrent related predicts current behaviour (Mertens, 2010) by determining whether scores on a specific test are systematically related to a criterion method collected at the same time as the test (McIntire & Miller, 2007). Concurrent validity finds its most important application when the evaluator has created a new measure that s/he believes is better than the previously validated one (Fink, (2005).

#### *Consequential validity*

Consequential validity refers to the social consequences of test interpretation and use (Mertens, 2010). Messick (1995) cautioned that this type of validity should not be viewed

in isolation as a separate type of validity, because it is integrally connected with construct validity. The researcher needs to identify evidence of negative and positive, intended and unintended, outcomes of test interpretation and use (Mertens, 2010). To ameliorate particularly the negative unintended outcomes, the test instrument should not miss something relevant or contain something irrelevant that interferes with the affected persons' demonstration of competence (Messick (1995).

The different validity evidence of performance assessment can be enhanced by removing an element of bias from the set tasks. Miller-Jones (1989) argues that “the use of ‘functionally equivalent’ tasks that are specific to the culture and instructional context of the individual being assessed” (p. 363) could be used to counter the problem of bias. This is because students' past experiences, their interests and the meaning they attach to the task are important factors not to be ignored.

Validity of inferences made of test results could be improved by increasing the number of assessment tasks or using a matrix sampling design, whereby different performance assessment tasks are administered to separate samples of students, during the design of performance assessment programme. Similarly, the extent to which a test's items actually represent the domain or universe to be measured is a very important factor in validating the test use (Moskal, & Leydens, 2000).

The design and development of tasks in collaboration with subject matter experts and stakeholders, particularly practitioners, is an important aspect in validating performance assessment (Burdett & Johnson, 2009). Subject matter experts complement each other in selecting appropriate items and by defining the content domain and universe in terms of both the subject matter and the type of behaviour or task desired from students (Mehrens & Lehman, 1991). Another way of validation could be through authentication of the marks on the final form by schools (Grima & Ventura, 1998). When teachers have not seen the development of the students' work over a period of time, teachers are asked not to authenticate the work. In that case, the candidates are called for an interview to establish whether or not the work was copied or recycled.

There is lack of agreement over how to validate analysis of qualitative research, and thus several contending positions (Lee & fielding, 2009). One example is through the quality of fieldwork, which addresses the adequacy of analysis by reference to factors such as the extent of fieldwork, effort devoted to coding, and the proportion of data accounted for by the most prominent analytic themes (Lee & fielding, 2009). On the other hand, there is validation through ethnographic authority (Hammersley & Atkinson, 1983), which gives credence to the researcher's interpretation since one would have witnessed events unfolding. Others propose that validity can be derived from systematic analytic procedures such as grounded theory (Glaser & Strauss, 1967) or micro-analysis applied to interview data (Agar & Hobbs, 1982). Lately, a postmodern approach to validating qualitative research through analysis which empowers research subjects (Altheide & Johnson, 1994) enjoys wide application. However, there is an increasing acceptance that what counts in establishing validity is the operation of the research community itself (Lee & fielding, 2009).

### 3.7.2 *Reliability*

Reliability is concerned with the consistency, stability and dependability of the scores. Popham (2005) contends that, as with validity, the reliability of the instrument is ascertained from the results obtained by administering the instrument. Such an instrument should be free of measurement error and ambiguity (Mertens, 2010), so as to obtain accurate measurements (Fink, 2005). For example, a self-administered questionnaire should be easy to understand, written in simple language at the level of the respondents, and have clear instructions. Reliability should therefore be calculated after every use because it is associated with the interpretation of the scores than with the instrument. Just like validity, there is no fixed reliability coefficient of an instrument (Mertens, 2010).

Reliability can be determined through several approaches. For instance, the *coefficient of stability (test-retest)* involves administering the same test to the same group of respondents on two occasions, with or without time lag (Mertens, 2010). The scores from both administrations are compared to determine the consistency of response. Aiken (1996) cautions that since the conditions of administration are likely to be different over



long time intervals from over short ones, the size of test-retest coefficients tend to be larger when retesting takes place after a shorter time than after several months. The test-retest method is appropriate only when test takers are not permanently changed by taking the test, or when the interval between the two administrations is long enough to prevent practice effects (McIntire & Miller, 2007). It is therefore important that whenever reporting the test-retest reliability, the length of time that elapsed between the two administrations should be stated. To circumvent the problem of practice effect, two parallel forms of the same test are given to the same test-takers.

The *Internal consistency* method was devised to overcome problems inherent in the repeated measures (Aiken, 1996; Mertens, 2010). In this type of reliability, a test is given only once to a group of respondents, split into halves before the set of individual test scores on the first half is compared with the set of individual test scores on the second half (McIntire & Miller, 2007). However, for this method to yield an accurate estimate of reliability, McIntire and Miller (2007) propose that the halves be equivalent in length and content. Questions are assigned to each half by random assignment, to balance errors in the score that can result from order effects, difficulty, and content. Since splitting the test shortens test length, hence decreasing reliability, Thorndike and Thorndike-Christ (2010) suggest adjusting the reliability coefficient using the Spearman-Brown formula.

An even better way to measure internal consistency is to compare individual scores on all possible ways of splitting the test into halves using KR-20 (McIntire & Miller, 2007). KR-20 is used to calculate internal consistency for testing whose questions can be scored as either right or wrong, while coefficient alpha is used to calculate internal consistency for questions that have more than two possible responses (Aiken, 1996; Thorndike & Thorndike-Christ, 2010). According to McIntire and Miller (2007), internal consistency is appropriate only for tests that are homogenous, that is those that measure one trait only.

*Scorer reliability* is concerned with how consistent the judgments of the scorers are (McIntire & Miller, 2007). When scoring requires making judgments, two or more scorers should score the test, using clear instructions for doing so. Scorer reliability can either be inter-rater or intra-rater reliability (Mertens, 2010), the former being concerned

with reliability between two independent raters, while the latter compares two data sets scored by the same rater. Score reliability can be expressed as either a reliability coefficient or expressed as a simple percentage of agreement between the two observational data (Mertens, 2010). Fink (2005) suggest that scorer reliability can be enhanced by training data collectors and providing them with guidelines for recording observations, monitoring and discussion of problems encountered by data collectors.

These different forms of determining the reliability of an instrument can be used to reduce the possibilities of lowering the reliability of the study. The discussion that follows outlines some of them.

As discussed above, employing multiple measurements of the same skill mitigates the problem of reliability (Airasian, 2005; Airasian & Russell, 2008; Crooks, 2004; Linn & Baker, 1996; Maxwell, 2008). Multiple rating could be done by different raters scoring students simultaneously, or the same rater scoring the same student at different times. Multiple raters can improve reliability (Rudner, 1994) because the errors of each observer tend to compensate for the errors of others (Thorndike & Thorndike-Christ, 2010). Multiple rating per se is not the solution to the problem of reliability as evidenced by various studies which have shown that written essays were scored differently by different raters and that even the same rater scored responses differently at different times (Rennert-Ariev, 2005).

The reliability of scoring through multiple rating can be enhanced if criteria are used (Nitko, 2004), and procedures put in place, such as recalibration of raters through refresher practice sessions (Johnson et al., 2009), to avoid 'rater drift' (Becker & Pomplun, 2006). Torrance (1995) and Shavelson et al. (1992) suggest training of observers on using scoring criteria.

Developing tasks of equivalent difficulties (Maxwell, 2004) which can be administered to different students also enhances reliability. Performance on one task provides a relatively weak basis of generalisation to other seemingly similar tasks. The limited degree of across-task generalisability in performance implies that performance needs to be assessed

across several tasks. It has been found that increasing the number of tasks is generally more important than increasing the number of raters, (Linn & Baker, 1996).

Because of the need to enhance reliability, some countries use highly standardised tasks and conditions, which have a tendency to reduce the validity of the tasks (James, 1994; Woods, 1991; Lennox, 2000). Others concentrate on aspects which are more readily measurable, such as knowledge and understanding. While this improves the reliability of measurement, it leads to the detrimental neglect of higher-level competencies and attitudes (Christofi, 1988).

The reliability of the results may be influenced by a number of factors, such as respondents' maturity, items that are ambiguous or unclear, and conditions of administration. Reliability of the instrument is improved by developing it in collaboration with stakeholders, thus ensuring that the construct to be measured is succinctly captured. The administration of a test is an important aspect in improving reliability (McIntire & Miller, 2007). Proper administration requires a manual detailing all procedures that all test takers should experience (Johnson et al., 2009).

In a situation in which qualitative data is collected, the researcher as the main qualitative data collection instrument should be sensitive, holistic, adaptable and responsive to changing circumstances, and observe activities silently (Guba & Lincoln, 1981) so as not to influence the outcome, thus improving the dependability of outcomes. Reliability is thus ensured by following an analytic inductive methodology in observation to test emergent propositions (Alder & Alder, 1994). Presentations of observational findings are then written in such a way that the accounts will contain a high degree of internal coherence, plausibility and correspondence to what readers recognise from their own experiences and from other realistic and factual texts (Alder & Alder, 1994).

However, the issue of validity and reliability should be approached with care, as Harlen (1994) points out that these are complementary terms which when one increases, the other becomes more difficult to attain.

### 3.8 CONCEPTUAL FRAMEWORK OF THE STUDY

The problem of Performance Assessment in senior secondary schools in Botswana, as discussed in Section 1.3, stemmed from lack of policy on continuous assessment resulting in variation in its conceptualisation. Teachers engaged in tasks of non-equivalent demands and non-standardised scoring resulting in the outcomes whose validity and reliability were uncertain. These necessitated an undertaking to understand and explore the characteristics and quality processes essential in the performance assessment of Agriculture Form Four students to ensure valid and reliable examinations in Botswana. Quality assurance processes should be embedded in the system if the outcome is to be reliable. Richard (1993) and Wild and Ramaswamy (2008) consider embedding quality into the processes as a process approach, whereby all the factors that have an impact on the students' achievement are examined. Such factors are found at both system-level and school-level. School-level factors are nested within the system implying that any improvement in the system results in the improvement in the school system.

Figure 3.1 presents factors affecting the validity and reliability of performance assessment. These factors draw from the work of Queensland Studies Authority (1998; 2008; 2009) and Wild & Ramaswamy (2008) in the case of policy formulation; Knostantopolous (2008), Knostantopolous and Chung (2009) for teacher workload; Jones, 2006 and Miller, Sen and Malley, 2007 and Finn et al., (2003) for student-teacher ratio; Stiggins (1997, 2002) and Wiggins (1998) for teacher training; Tindal and Hladyna (2002) and Nitko and Brookhart (2007) for resources provision; Mamary (2007) for school leadership and monitoring and supervision; McMillan (2004) and Popham (2005) for learning autonomy; Deakin Crick, Broadfoot and Claxton (2002) and Harlen (2006) for student motivation; Airasian and Russell (2008) and Nitko (2004) for multiple modes of assessment and multiple rating; and Wiggins (1998) for student readiness for assessment. It has at its centre performance assessment, which is influenced by the both system-level factors and school-level factors. System-level factors include, but are not limited to, assessment policy, monitoring and supervision, student/teacher ratio, teacher training, teacher workload, and provision of resources. On the other hand, school-level factors include school leadership; learning autonomy, student motivation, multiple modes of

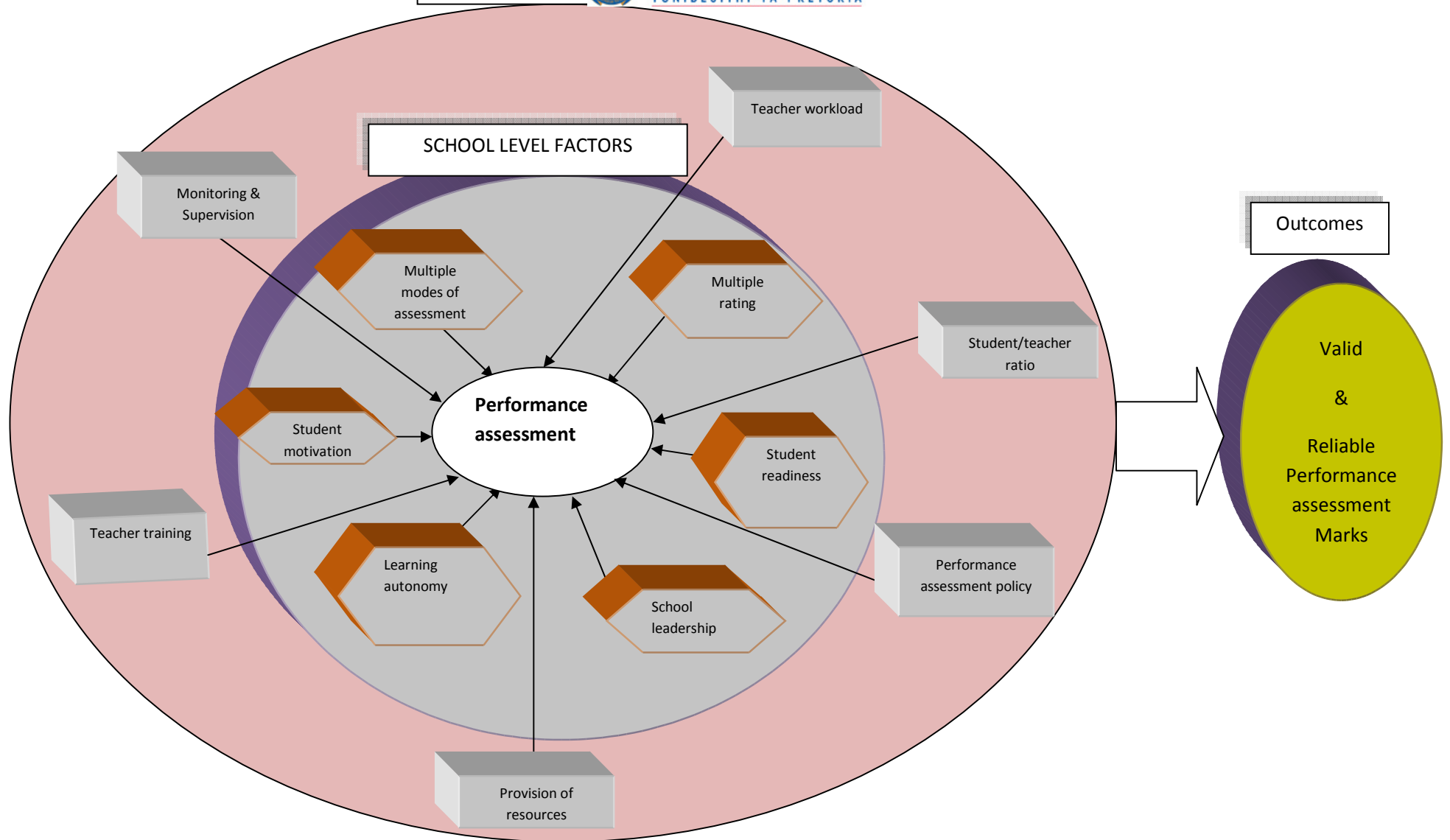


Figure 3.1: Factors affecting the validity and reliability of performance assessment marks

assessment, multiple rating, and student readiness.

### **3.8.1 System-Level Factors**

These are factors that are determined at ministerial level and in most cases a top-down approach is followed. Schools and teachers do not have much say in decision-making. They are required to implement what has been decided for them.

#### *Assessment Policy*

Implementing Performance Assessment has to be guided by a policy to produce valid and reliable performance marks for certification (Wild & Ramaswamy, 2008). The policy should outline among other things approval or accreditation based on the quality of teachers to conduct performance assessment, physical and material resources such as tools and livestock. The policy should also spell out objectives that should form performance assessment tasks, who should assess, how many tasks to be done, the roles of the teacher, students, supervisors, and how quality is to be assured (Queensland Studies Authority, 1998; 2008; 2009).

#### *Provision of resources*

Implementing performance assessment on large-scale requires massive resources which are costly (Tindal & Haladyna, 2002), just like standardised testing does. Resources such as garden, laboratory, laboratory equipments, tools and equipments, exemplar tasks and assessment materials, and time play a significant role in students' achievement. However, it should be noted that, the presence of equipment and other learning materials do not necessarily imply effective learning and assessment. Performance Assessment by nature requires a lot of time. For example, a portfolio might be developed over a year (Mills, 1996; Johnson et al, 2009). With many students in a class, this might present a mammoth task for the teacher since scoring performance assessment is also a difficult and often time-consuming activity (Nitko & Brookhart, 2007).

#### *Teacher training*

Training teachers in assessment methods is vital for successful implementation of performance assessment (Richard, 1993). Teachers are at the forefront of classroom assessment but it was noted by Stiggins (1997) that, "A lot of people involved in education, including teachers do not understand how assessment should be done and why it is done" (p.

2). Training in assessment has never been a prominent part of teacher training (Stiggins, 1997, 2002; Stiggins & Conklin, 1992; Wiggins, 1998) rendering teachers to be unprepared to assess their pupils especially on performance tasks (Kellaghan & Greaney, 2003). For example, Stiggins (2002) noted that “only about fourteen out of fifty states in America require that pre-service teachers take an assessment course. “Only three states require competence in assessment as a requirement for being licensed as a principal, and no state certifies that competence” p. 21.

Despite that, Stiggins (1997) observed that teachers spend most of their time engaged in assessment activities, and yet teacher classroom assessment is weighed proportionally little to the overall mark whenever it is used for summative purposes. However, if teachers are properly trained and given enough support resources they can design and develop sound assessments (Maxwell, 2004). Once teachers have acquired the necessary expertise, they can act professionally and ethically and typically take up the challenges when they are given the responsibility (Maxwell, 2004).

Teachers who lack training in assessment are apt to approach formative assessment in an essentially behaviourist approach. The assessment converges on the assessor’s agenda – of trying to find out whether the child knows, understands or can do predetermined things, rather than divergent assessment which emphasises learners understanding (Torrance & Pryor, 1998) and to provide feedback to students on how they have performed in certain objectives and what else they might need to do in order to realize incremental improvement. If we need incremental continuous improvement (Goetsch & Davis, 1997), performance assessment should occupy the centre stage in pre-service and in-service teacher training programmes.

Though in-service training is normally provided, most of the time, initiatives are poorly conceptualized, and insensitive to the concerns of individual participants. Halsall (1998) refers to such designed training programmes as quick fix solutions to the schools’ problems. Properly trained teachers in assessment should be able to engage in self-assessment which makes them aware of their own limitations, and those of the techniques they use. If teachers lack competence in some areas of assessment, they should not engage in assessment, no matter how much they are persuaded by the school officials (Salvia & Ysseldyke, 1998). Students too, according to Wiggins (1998) can be given the same training that the assessors receive for them to be able to judge that their work is not up to standards.

Lack of training in assessment causes teachers to unwittingly misinterpret performance assessment and deemphasise or neglect untested material whenever they are engaged in assessing it (Tindal & Haladyna, 2002). Some view school-based assessment (Torrance, 1995) as an extra workload (additional marking, record keeping, and so on) leading to teachers inflating students' marks particularly if they are to be used for summative purposes (Raivoce & Pongi, (2000).

#### *Supervision and monitoring*

Monitoring and supervision is extremely important for the success of any project. Supervision must not only be viewed in terms of finding faults on the teacher, but rather as a continuous process aimed at improving teacher performance hence improvement in students learning (Mamary, 2007). Monitoring of school-based performance assessment should be done on daily basis by the senior teacher and routinely by administration. External monitoring is also essential to ensure that teachers do not deviate from standards.

#### *Teacher Workload*

Workload normally is positively correlated with class size. The more the students are in a class, the more the work for individualised assistance in a student-centred approach (Knostantopolous, 2008). Workload is probably one of the reasons why teachers ultimately adopt a student-centred approach to instruction and assessment, despite their consciousness of the little impact it has on students' learning of high-order thinking and abstract reasoning. Workload is increased by a lot of recording involved in performance assessment of the student achievement and progress which majority of teachers find it a nightmare (Knostantopolous and Chung, 2009; Torrance & Pryor, 1998). Performance assessment taking place at school level to be included in the certification exerts extra work load on both teachers and students (Fung et al., 1998), because timetabling does not cater for it (Abram, 2008).

#### *Student-teacher ratio*

In developed countries, the student/teacher ratio is very small, facilitating individualised instruction and assessment. For example, the G-8 countries' class size ranges from 10 in the Russian Federation to 16 in the United States of America at secondary level (Miller, Sen & Malley, 2007). Jones (2006) reported that in order to increase the connection between materials taught and what students experience in the field setting, the class sizes needs to be



25 or less. If there are just too many students in the classroom, the teacher's assessment focus tends to be on class, or perhaps the small group, rather than the individual student.

Finn et al. (2003) identified reasons as to why and how small classes yield better results. These include: more participation, engagement and identification. There is also more teacher time per student for diagnosing learning problems, working with portfolios, correcting homework, reading with each child and more time for individualised assessment and increased time on task. However, Patchen (2004) and Wiles and Bondi (2000) posit that these can only be effective if teachers change their teaching styles.

### ***3.8.2 School -Level Factors***

These are the factors that schools can vary to suit their needs. They include leadership, learning autonomy, student readiness for assessment, multiple modes of assessment, multiple rating, and student motivation.

#### *School leadership*

The school head has the duty to manage testing and assessment for the effective running of the school (Mamary, 2007). This includes appropriate conduct of performance assessment which can only be effectively implemented if school management is committed to the responsibility of quality assurance. One of the major functions of management is the formulation, implementation and review of a quality policy (Richard, 1993). If testing and assessment is thoroughly monitored, school policy decisions and instructional leadership/support formulated based on information obtained from quality standard tasks may lead to student achievement hence improved school functional system (Stiggins, 1997).

Supervision is no longer confined to lesson observation (Mamary (2007), but supervisors need to work with teachers on continuous basis and create a school climate in which teachers' self-assessment and co-assessment becomes the culture with the aim to succeed academically. The supervisor's role is to link the purpose and goals of the school to the role of the supervisee and to the improved assessment of the students.

#### *Learning autonomy*

The teacher-centred approaches permeating classroom instruction uses direct instruction to whole classes, and appears most applicable: to a well-structured body of knowledge where

skills do not follow explicit steps, in introducing and explaining new concepts, in showing how specific pieces of information fit into logical structures and in reviewing and summarising information (McMillan, 2004; Popham, 2005). The need to allow students to actively construct their own knowledge through active participation heralds a paradigm shift. In student-centred didactic, the teacher's role is delegated to explaining basic concepts and skills in facilitating group learning. Recent student-centred learning approaches are based on instructional strategies such as cooperative learning, problem-based learning, discussion, discovery learning, and collaboration. These instructional strategies are feasible in an environment where class sizes are manageable, teachers having skills in assessment, and viable contextual factors promoting effective classroom ecology.

### *Student motivation*

Students should be prepared to continue learning after school. This can only happen if they are motivated to learn (Deakin Crick, et al, 2002). All students want and have the capacity to learn (Greenwood & Gaunt, 1994). The aim of learning is to continually improve their performance and self-esteem, not to measure their failure. Motivation for learning can be fostered in the form of interest, goal orientation, locus of control, self-esteem and self-efficacy, and self-regulation (Harlen, 2006). Motivated students with learning goals have the following characteristics (Dweck, cited in Torrance & Pryor, 1998, p. 85).

- choose challenging tasks regardless of whether they think they have high or low ability relative to other children,
- optimise their chances of success,
- tend to have an incremental theory of intelligence,
- go more directly to generating possible strategies for mastering the task,
- attribute difficulty to unstable factors e.g. insufficient effort, even if they perceive themselves as having low ability,
- persist in their endeavour, and
- remain relatively unaffected by failure in terms of self esteem.

A transparent assessment system where students and the teacher consult each other about assessment; developing rubrics jointly, and applying rubrics to common examples of student work and then discuss the results (Stiggins, 1997; Mergendoller, Markham, Ravitz, & Larmer, 2006), motivates students to achieve.

### *Multiple modes of assessment*

Assessment in educational settings is a multifaceted process, encompassing the way students perform a task in a variety of contexts or settings (Airasian, 2005; Mamary, 2007). As such there are different kinds of achievement to assess which include knowledge, skills, product, reasoning, and dispositional (Stiggins, 1997). Various assessment methods have to be repeatedly employed to reflect those achievements and to allow for all the intended learning outcomes to be appropriately assessed (Maxwell, 2004). Tindall and Marston (1990) identified three sources of information which differ greatly in the type of data and methods employed. These are observations which are non-interactive, interviews which are interpersonal, and testing which focuses on quantifying performance.

### *Multiple rating*

As discussed above, multiple observation of students' performance provides more reliable and accurate information (Airasian, 2005; Airasian & Russell, 2008). Multiple rating could be done by different raters scoring students simultaneously or the same rater scoring the same student at different times. Raters can score the same piece of work differently and even the same rater can score it differently at different times (Rennert-Ariev, 2005). It is argued that multiple raters can improve reliability of performance assessment just as multiple test items can improve the reliability of standardised tests (Rudner, 1994). The reliability of scoring through multiple rating can be enhanced if the criteria are used (Nitko, 2004), and procedures put in place such as recalibration of raters through refresher practice sessions (Johnson et al, 2009), to avoid rater drift (Becker & Pomplun, 2006).

### *Student readiness for assessment*

The assessment of students is a social act that has social and educational consequences. Students should therefore have the right to the assessment procedures and should have the willingness to complete the assessment. Emphasising the need for student readiness for assessment, Grant Wiggins (1998) had this to say:

Gone are days when silent examinees sitting in rows, answering uniform questions with orthodox answers in blue books or on answer sheets with No. 2 pencils. Gone are arbitrary calendars that dictate that students must all be examined simultaneously, regardless of readiness (p. 3).

Data generated through assessment is used to make decisions about the students, and the decisions could significantly adversely affect an individual's life opportunities if assessment is improperly made. Salvia and Ysseldyke (1998) assert that those who accept students must accept responsibility for the consequences of their work, and they must make every effort to be certain that their services are used appropriately. Assessment should not be viewed as a means to rank order students to select those who can proceed for further education or employment, but rather to impart social and life skills that they can use outside school.

### 3.9 CONCLUSION

Performance assessment is being reintroduced by a number of countries as a reform measure in assessment (Khoo & Idrus, 2004; Maughan, 2004), because of its ability to improve learning on the one hand, and its complete evaluation of the student's capabilities on the other hand. Performance assessment includes products and performances such as portfolios, projects, and experiments which when properly implemented results in the acquisition of complex thinking skills, problem-solving skills, and abstract reasoning (ARG, 2006; Shepherd 2000, 2008). These skills give the kind of challenges, diversity and flexibility that make assessment more realistic and educative (Wiggins, 1998) because the thinking processes that students undergo to construct responses is assessed rather than simply auditing learning (Airasian, 2005). Assessing the thinking processes helps students improve their learning (ARG (2002), resulting in deeper and critical thinking.

Performance assessment is engaging, open and uses criteria which describe the expected quality of performance students must meet (MacMillan, 2000). Feedback is provided to assist students to improve and attain the expected quality of performance. The authenticity of performance assessment allows students to perform in the context of the real-world situations in which the skills are to be applied (Johnson et al., 2009; Nitko & Brookhart, 2007). Whenever performance in real-world is not plausible, simulation is used (McMillan, 2004). Mistakes are made during the trial period and they help students grow and develop over time. Students' growth is influenced by many factors and cannot be assumed to be the same (Neill & Medina, 1992). Though clearly defined criteria are used, an element of subjectivity can creep in when assessment is carried out by one assessor (Thorndike & Thorndike-Christ,

2010). Quite often, performance assessment is used to assess those skills which cannot be assessed by paper-and-pencil which the teacher is best placed to do.

The control of performance assessment by the external bodies varies across countries. Some countries have complete control from determining the curriculum, developing task frames that guide the development of tasks, through developing standardised performance materials, standardising administration, centralise marking to manipulation of the outcomes through statistical moderation (Berry, 2008; Broadfoot, 1994; Lennox, 2000; Singh, 2004). In such countries, performance assessment is used to supplement paper-and-pencil examinations with the consequences of low weightage attached (Kanjee & Sayed, 2008; Maughan, 2004; Singh, 2004; Van der Berg & Shepherd, 2010).

In some countries, performance assessment has completely replaced one-off final examinations (Broadfoot, 1994; Gasemann, 1993; Queensland Studies Authority, 1998). The development of assessment tasks and procedures is the responsibility of the teacher, and the resulting outcome involves an interpretation of the final product of the student's work by a judgement of the standard it demonstrates when compared to a set of grade descriptors (Mercurio, 2008; Maxwell, 2004).

Moderation of the assessment throughout its conduct is the way to ensuring valid and reliable outcome of performance assessment. Subjecting marks to statistical moderation to normalise them when little is known about how the marks were produced does not help in improving the quality of assessment. As indicated above, this results in low weightage being attached to such assessment mode which has the adverse effect of lowering teachers' morale. Since teachers spend more time in this kind of assessment, they expect a corresponding high weight to be given to it. Emphasising training of teachers on performance assessment can result in more valid marks being produced hence increasing the weight of performance assessment to the final mark.

Performance assessment by nature is valid because it represents the actual activities in real life. Properly crafted performance tasks can therefore rate highly in validity when guiding principles are followed during their construction and scoring. Since a score from a single assessment is not reliable, performance scores are generated over a period of time, using different raters. Reliability of performance assessment is low particularly when the stakes are

high and teachers tend to collude with both students and parents to inflate students' marks so that they can pass examinations.

Meta-analysis of a number of countries on performance assessment implementation has revealed similarities and differences, particularly with regards to quality assurance procedures. Performance assessment in developed countries emphasises quality assurance procedures, whereas African countries seem to lag behind in emphasising quality assurance aspects due to high level of costs and administrative complexities associated with performance assessment. A number of African countries are still battling with access to education hence quality assurance is given secondary treatment. Thus, conceptualisation of performance assessment followed by entrenching quality into the system is of paramount importance for its success.

Botswana as an African country has to overcome the same problems of performance assessment implementation and entrenching quality assurance processes in the system. Currently emphasis is on quality control carried out once at the end of the year, either by visiting moderators who go out to schools to dictate what teachers should do or applying statistical moderation to the teacher marks which are not necessarily valid. There is no literature on how quality assurance is assured in performance assessment in Agriculture in Botswana schools. Anecdotal evidence suggests that there is quality control at the end of the year in Agriculture performance assessment by moderators. This study will take a step forward to understand and explore the characteristics and quality assurance processes needed in performance assessment of Agriculture Form Four students, and develop quality standard task and assessment materials for use in a quality embedded environment.

## CHAPTER FOUR

### RESEARCH DESIGN AND METHODS

#### 4.1 INTRODUCTION

This chapter discusses the research design and methodology of the overall study, to understand and explore the characteristics and quality processes needed in the performance assessment of Agriculture Form Four students to ensure valid and reliable examinations in Botswana. The overall design for the study is the design research in which the context of the problem is first understood by conducting a survey and then developing an appropriate intervention. The two main research questions are:

- 1. How valid and reliable are the performance assessment processes in Botswana schools?*
- 2. How can quality assurance processes be developed in order to produce valid and reliable marks for BGCSE Agriculture performance assessment?*

In this chapter, an overview of the research design is presented. Detailed designs are presented in chapters six and seven. In Section 4.2, the paradigm underlying knowledge base of this study is outlined. Section 4.3 presents the overview of the research design for the study. Sections 4.4 and 4.5 delineate the research design for baseline survey and the intervention respectively. Methodological norms are outlined in Section 4.6 while Section 4.7 concludes the chapter. A baseline survey will be conducted to understand the context of the problem and its findings will form the basis of intervention development. The research design and methods for baseline survey and intervention development are discussed separately in this study.

#### 4.2 PARADIGM UNDERLYING THIS STUDY

The need to link the research process to a philosophical paradigm is the subject of ongoing debate, with some advocating for its application (Schwandt, 2000; Ladson-Billings, 2005; Mertens, 2010), and others against (Patton, 2002). Despite the debate, there is an overwhelming agreement that philosophical paradigms serve the purpose of providing a framework for discussion (Cresswell, 2009; Guba & Lincoln, 1994; Lather, 1992; Mertens, 2005; Morgan, 2007; Tashakkori & Teddlie, 2003). Although many researchers apply two or

three paradigms (Schwandt, 2000) in the same study, and sometimes that might not even be adequate basis for design (Burkhardt, 2007).

The paradigm pertaining to this study is based on pragmatism (Driscoll, 2000; Mertens, 2010), which provides an underlying framework for mixed methods research. Since social science inquiry is unable to access the “truth” about the real world solely by virtue of a single scientific method (Morgan, 2007), mixed methods research is often applied. In mixed methods, both qualitative and quantitative methods are applied to optimise their strength of complimenting each other (Onwuegbuzie & Johnson, 2004). This results in answering research questions more appropriately than limiting to methods that belong to either of the two.

Pragmatism is discussed in terms of the philosophical assumptions of ontology, epistemology, axiology and methodology. Ontology is the assumptions about the nature of what exists and what is viewed as reality. Epistemology is the assumptions about the nature of knowledge deemed appropriate within the value system. Axiology is the assumptions about ethics and beliefs. Methodology is the assumptions about what works best for acquiring knowledge (Mertens, 2010).

In pragmatism, the ethical goal of research is to gain knowledge in the pursuit of desired ends (Mertens, 2010). That is, all that is worth valuing is a function of its consequences (Christians, 2005). In this regard, Mertens (2010) observed that the belief systems of pragmatism are closely aligned to those of the constructionists. The ontological position is the search for contextual multiple truths rather than objective truth (Ornstein & Hunkins, 1993; Donald, Lazarus & Lolwana, 2002; Mertens, 2010) as individuals have their own unique interpretation of the world. Pragmatists believe that knowledge can be created “through behaviour where different people or groups of people come together for a common cause (Mertens, 2010). The results that can provide solutions to social problems are most important than the “laws and rules governing what is recognizes as the truth” (Mertens, 2010, p.36-37). What works more efficiently for the current situation is what is valued than something proven to work in other contexts.

The epistemology in pragmatism is that knowledge can be ascertained by means of reason or experience, but it is always provisional (Tashakkori & Teddlie, 2003). Pragmatists hold absolute knowledge as a worthy but probably unreachable goal, thus they emphasize theories



of meaning of what works, with the understanding that this may not reflect reality, assuming that reality is acknowledged but not presumed to be known directly (Morgan, 2007; Teddlie & Tashakkori, 2009; Mertens, 2010). Knowledge was considered a transaction between learner and environment (Butler-Krisber, 2010; Ornstein & Hunkins, 1993), both of which were as a consequence of constantly changing transactions or experiences (Ornstein & Hunkins, 1993). This is contrary to post-positivist thinkers whose nature of knowledge is grounded in the quantitative methods that aim to establish the objective truth which can be generalised to the entire universe (Mertens, 2010).

To establish knowledge that can bring positive consequences to performance assessment in schools, a mixed method approach was used guided by the purpose of research which sought to determine the validity and reliability of performance assessment processes in Botswana schools, and then developing quality assurance processes in order to produce valid and reliable marks for BGCSE Agriculture performance assessment. Mixed methods offer a practical solution in that a quantitative approach can be used as a baseline or exploratory means for identifying actions that a certain group of people adopt. A follow up with qualitative approaches helps understanding their actions through thick descriptions, or vice versa. The methods are not cast in stone, and they can be varied or modified to suit “the community that serves as the researcher’s reference group” (Mertens, 2010, p. 38).

Students were required to explore rather than explain during their practical lessons, hence methods of learning by doing to solve problems were emphasised rather than mastering organised subject matter, bearing in mind that students learn in different styles and at different rates, as discussed in Section 3.3. Learning was considered, according to the scientific method, a process of reconstructing experience individually or in a group, to solve problems that vary in response to the changing world and climatic conditions that directly affect agricultural activities. The adoption of such an approach to learning was to prepare the student for the future, as per the standards of the curriculum discussed in Section 2.5.3.

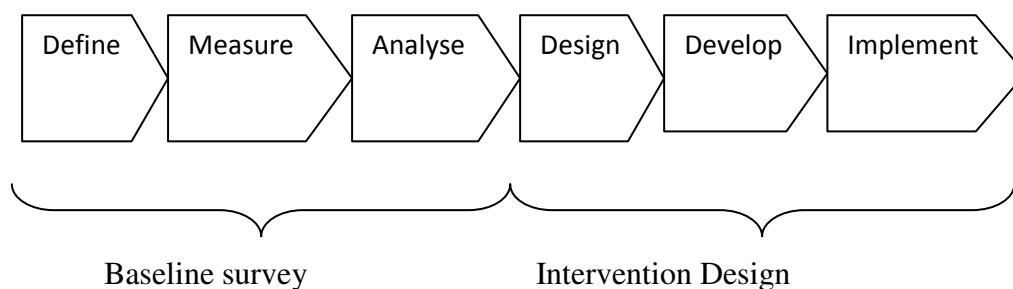
The researcher’s goal in this study was to understand the multiple social constructions of meaning and knowledge in agriculture performance assessment achieved by creating knowledge through design-based research, in which the researcher collaborated with stakeholders in iteratively developing prototypes of the intervention. Effectiveness of the intervention was ascertained by its ability to practically solve the specific problems encountered in performance assessment in agriculture (Mertens, 2010), rather than

conforming to the ‘true’ condition in the real world (Mertens, 2010, p.36). Applying mixed methods and triangulation strategies in data collection, such as questionnaires, interviews, observation, and document analysis helped the researcher to know how teachers conducted performance assessment, why they did it the that way, how it could be done differently to yield better results applicable to their context. Thus, the interest was in methods that yielded results (Maxcy, 2003).

### 4.3 OVERVIEW OF RESEARCH DESIGN

This study employed the research design which was conducted in two phases addressing the two main research questions discussed in Sections 1.6 and 4.1. During the first phase of the study, a baseline survey was conducted to address the main research question 1, which sought to understand the conduct of performance assessment in Botswana schools. A baseline survey was suitable in this particular case to collect information quickly by asking respondents to complete a self-administered questionnaire and administering an interview.

The baseline survey identified the problem context (Barab & Squire, 2004; Colton & Covert, 2007; Kelly, 2004), and described assessment practices and processes, as well as points of views and attitudes held by practitioners (Cohen & Manion, 2000; Persse, 2006). These were compared to policies and procedures for both nationally and internationally community. The successful conduct of baseline survey was achieved by infusing the first steps of DMADDI methodology of Design For Six Sigma (DFSS) approach, shown in Figure 4.1 (below).



*Figure 4.1: The DMADDI approach of DFSS (Source: Islam, 2006. p.52)*

DMADDI is the acronym for the six steps of the DFSS approach (Define, Measure, Analyse, Design, Develop and Implement) (Islam, 2006). DFSS is a process-design approach which considers the system as a whole (Abramowich, 2005; Oakland, 2003; Persse, 2006), ensuring

that the processes of the highest quality and reliability are designed (Breakthrough Management Group, 2007; Goetsch & Davis, 1994; Rainey, 2005). Its aim is to eliminate defects from existing processes and products or services (Islam, 2006).

In the second phase, the needs of the users together with design specifications (Abramowich, 2005) formed the inputs for prototype development to produce a better intervention to address the problem. The design of the intervention employed the design-based approach, superimposing the last three steps of DMADDI, namely design, develop and implement, on the modified design of Mafumiko (2006), as shown in Figure 4.2 (below). Although the initial intention was to develop five prototypes, it was not possible due to disruptions towards examinations time in schools (*Mmegi* Newspaper, 26 October, 2009.p.4), which affected the review by experts. The intention of the expert review was mainly to review and outline resources that were needed for effective implementation of the intervention. Consequently, four prototypes of the standard task and assessment materials were developed.

The development of prototypes was carried out in collaboration with stakeholders and practitioners (Collins, Joseph & Bielaczyc, 2004; Gravemeijer, 1998; Hoadley, 2002; The Design-Based Collective, 2003), to increase adoption. The development of each prototype is fully discussed in Chapters Six and Seven. Formative evaluation was an integral part of the intervention development and feedback was incorporated into the redesign to improve successive prototypes. Design research was considered appropriate for this study because of the flexibility in developing an intervention stage by stage within the problem of the context, facilitating understanding of implementation problems that practitioners experienced.

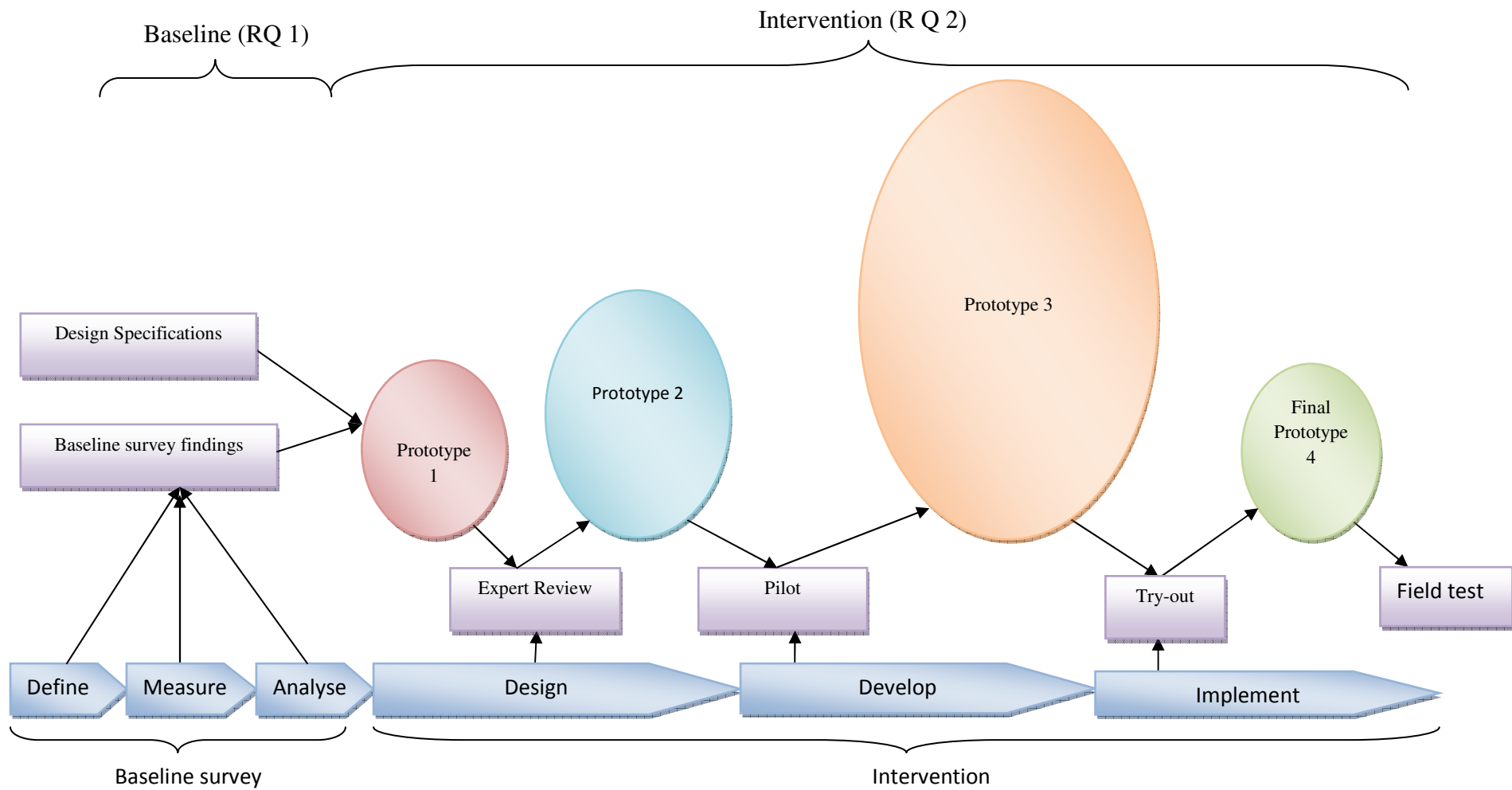


Figure 4.2: Research design (Source: Adapted and modified from Mafumiko, 2006, p.48)

#### 4.4 RESEARCH DESIGN FOR BASELINE SURVEY: PHASE ONE

The various phases of the research design are: research design outlined in Subsection 4.4.1, research methods presented in Subsection 4.4.2, and data analysis discussed in Subsection 4.4.3. The section on the research methods is further subdivided into sample and participants, instrument development and data collection strategies, and data collection procedures.

##### ***4.4.1 Research design***

As outlined in Section 4.1, the main research question in phase one was to determine the validity and reliability of performance assessment processes in Botswana schools through a baseline survey research. The baseline survey explored the participants' conduct and understanding of the performance assessment practices in relation to both national and international policies and procedures (Breakthrough Management Group, 2007). The baseline survey focused on two of the five regions in Botswana. The nature of regions is described in Subsection 4.4.2. To address the main research question, it was operationalised into three sub-research questions (a), (b) and (c) discussed under Sections 1.6 and 4.1.

##### ***4.4.2 Research methods***

The sample and participants, instruments development and data collection strategies, the research procedure, and data analysis constitute research methods which are explained in detail in the following sub sections.

###### *4.4.2.1 Sample and Participants*

Schools in Botswana fall under five regions. Two regions closer to the researcher were purposively sampled (Fink, 2005; Wiersma & Jurs, 2005), and all agriculture teachers in these two regions which had a total of thirteen schools formed the sample. Purposive sampling was employed to reduce expenses and time associated with travelling and accommodation. The two regions although purposively sampled were considered typical of all government schools (See Subsection 2.5.3 and Section 2.6). This is because funding of the schools is centralised and there is fair distribution of resources such as infrastructure, tools and equipments, irrespective of the school's geographical location. Teachers are also deployed centrally by the ministry headquarters and have no choice of schools to be posted to. In addition, there is incentive in the form of money and accelerated promotion to attract

teachers to go to remote areas. Schools offer the same curriculum which is produced and examined centrally at the end of senior school.

The participants included Agriculture teachers and school administrators<sup>7</sup>. All Agriculture teachers and school administrators in the sampled schools completed the questionnaires that had been piloted in two schools from different regions. Table 4.1 (below) outlines the number of participants for both the pilot and the final survey.

*Table 4.1: Sample of participants in the study*

Participants	Number of participants in	
	Pilot	Main Data
Teachers	14	68
Schools	2	13
Senior Teachers	2	13
School administrators	4	26

#### *4.4.2.2 Instrument Development and data collection strategies*

As a first step in constructing the instrument, the researcher identified the constructs to focus on, from both literature search and through discussion with relevant authorities. This facilitated the acquisition of information to answer the research questions, hence providing accurate and useful information for decision-making (Colton & Covert, 2007). Similar instruments were examined at this stage to see how others had measured related constructs. No instrument was found that could effectively measure the constructs under investigation, hence instruments were developed borrowing some items from other sources such as Januario's (2008) instruments, which were used as they were or modified to match the context (questions 13, 14 & 15 in the Teacher's questionnaire). A matrix presented in Appendix 4.1 illustrates the constructs and questions addressing them.

To enrich in-depth understanding of the assessment processes, various documents related to assessment were evaluated, such as teaching syllabuses, assessment procedures, and colleges' assessment course outlines. Relevant authorities on the subject evaluated the instruments

<sup>7</sup> For the purposes of this study, school administrators were considered to be school head and deputy school head

before they were administered and based on the feedback, instruments were revised in preparation for piloting.

**Questionnaires** for both teachers and school administrators were developed for the baseline survey. They were preferred over other data collection instruments because they collected data within a short period of time, as well as providing advance information to focus the interviews (Colton & Covert, 2007; Mindes, 2007).

**Teacher questionnaire** consisted of closed-ended questions seeking respondents' background information, such as sex, age, length of teaching, post of responsibility, marking experience and qualification. Different scales to measure the feelings, attitudes, opinions and frequencies of occurrence (Alreck & Settle, 1995; Dick, Carey & Carey, 2001; Dyer, 1995) were constructed. The choice of scales was based on their simplicity, flexibility, and economy. Also of significant importance was their ability to give an overall index of a construct, as well as individual weighting of items (Alreck & Settle, 1994; Colton & Covert, 2007). The instrument comprised the following scales: modes of assessment; learning autonomy; assessment for learning; availability of resources; monitoring and supervision; standardising of marking; and attitudes towards performance assessment. The instruments consisted of open-ended questions to capture thick descriptions of the respondents' opinions (Goddard III & Villanova, 2006), (See Appendix 4. 2).

**School Administration questionnaire** like the teachers questionnaire, contained both closed-ended questions seeking the respondents' background information, and three scales seeking the respondents' perspectives on assessment practices; resources; and monitoring and supervision. Some open-ended questions seeking the respondents' opinions on the issue were also included, (See Appendix 4.3).

**Teacher interview schedule** was directed by questions that sought to find out how performance assessment was conducted in schools, with particular reference to resources, how the administration assisted teachers during the conduct of performance assessment, and the challenges they encountered during the conduct of performance assessment. The interview schedule is presented in appendix 4.4.

#### *4.4.2.3 Data Collection procedure*

The research was preceded by a literature search to gain insight on how to design and develop tasks and their implementation. This search was extensive to encompass an international perspective, before narrowing it to Africa and then Botswana. Emphasis was placed on understanding how quality was assured in formative performance assessment included in determining the final grade of the student. The depth and breadth of the literature review informed the conceptual framework, which provided direction for the study.

The extent of the problem in performance assessment was determined by the baseline survey, the initial stages of which involved assembling a team of practitioners and stakeholders to identify and define the problem, followed by development of self-administered questionnaires and interview schedule (See Section 4.4.2.2). Triangulating of data sources was aimed at improving the validity and reliability of information collected (Mertens, 2010). The constructed instruments were edited by four subject content specialists, particularly for content validity, logical sequencing of items and questions comprehensiveness. Two language experts also vetted the instruments for correct language and grammar usage. Nine measurement specialists finally subjected the instruments to verification for psychometric soundness.

After incorporating suggestions from editors and verifiers, instruments were piloted in two schools in February 2010, and involved fourteen teachers and four administrators (See Table 4.1, above). Further modifications on the instruments based on piloting outcomes helped to remove ambiguity of items to elicit appropriate responses. For example, the time for administering was found to be too long, some items were found to be ambiguous and reworded, and the language of some items had to be revised.

For collection of the main data, the researcher delivered self-administered questionnaires (discussed in Subsection 4.4.2.2) to thirteen schools in the two nearby regions, and after a week the researcher returned to collect. However, majority of school administrators had not completed. A reminder was sent three weeks after the first issuance. The return rate improved, as indicated in Table 4.2 (below). Interviews were conducted with nine teachers, among them two senior teachers who were conveniently sampled from each of the two regions to identify information-rich participants for in-depth understanding of the



phenomenon (Mertens, 2010). Data was then subjected to analysis, as discussed in the following section.

*Table 4.2: Administrators' and teachers' response rate to the questionnaires*

<b>Respondents</b>	<b>Sampled respondents</b>	<b>First return rate</b>	<b>Second return rate</b>
School Administrators	26	12	21
Teachers	68	57	57

#### **4.4.3 Data Analysis**

Preliminary data screening was conducted using SPSS to test for adequacy of factor analysis to be analysed. Screening included determination of the correlation matrix coefficients, Cronbach's alpha, multi-collinearity or singularity, and sampling adequacy. The Kaiser-Meyer-Olkin (KMO) value of 0.7 was considered adequate for principal components analysis to be carried out. Bartlett's test of sphericity for testing the null hypothesis should be significant (indicating sufficient correlations between the variables) (Field, 2000; Meyers, Gamst & Guarin, 2006; Pearson, 2010; Tabachnik & Fidell, 2001).

Exploratory factor analysis using a principal component extracting method and varimax rotation was then computed. Factors were extracted using the Kaiser-Guttman retention criterion of eigenvalues greater than 1.0, as well as the scree plot to provide the best solution (Kremelberg, 2011; Meyers et al., 2006). In practice, Tabachnick and Fidell (2001) argue that a robust solution should account for at least 50% of the variance. The 0.40 cut-off point was used for factor loadings (Dancy & Reidy, 2002; Field, 2000; Hair, Anderson, Tatham & Black, 1995; Stevens 1992) because it represented substantive values. Scores produced through factor analysis were used for further analysis, such as t-test and ANOVA.

Data from open-ended items were analysed qualitatively into themes, using thick descriptions to capture respondents' views (Butler-Kisber, 2010; Wiggins & Riley, 2010). All interviews were transcribed in full and an iterative process of qualitative analysis was employed (Coolican, 2006), drawing upon elements of grounded theory (Butler-Kisber, 2010). Analysis began with open coding, assigning codes using respondents' own words as far as possible. A

set of categories were then constructed which were thought to best describe interviewee's conceptions. These categories are outlined in the next section.

#### 4.5 RESEARCH DESIGN FOR THE INTERVENTION STUDY: PHASE TWO

In Section 4.4, a baseline survey was discussed which formed phase one of this study aimed at understanding the validity and reliability of performance assessment processes in agriculture in Botswana schools. Based on the understanding of the processes prevailing, phase two of the study aimed at developing an intervention that infused quality assurance processes to produce valid and reliable marks.

As discussed in Section 1.3 in Chapter One, the major problem in performance assessment in Agriculture was the absence of quality standard tasks and assessment materials to guide teachers. The development of the intervention applied design-based research, discussed in detail in Sections 4.5.1 to 5.5.5, guided by the following research question, which was divided into sub-questions as discussed in Sections 1.6 and 4.1.

*How can quality assurance processes be developed in order to produce valid and reliable marks for BGCSE Agriculture performance assessment?*

Design specification outlined in Section 6.2 guided the iterative development of the intervention (standard tasks and assessment materials), employing a mixed method approach (Maxcy, 2003) which used multiple approaches in answering research questions (Johnson & Onwuegbuzie, 2004). The development was interspersed with formative evaluation after successive enactment. Feedback from the evaluation process was factored into the redesign of the subsequent prototypes until the final prototype was field tested.. The ultimate goal was to characterise the design elements (Gravemeijer & Cobb, 2006; Plomp & Nieveen, 2007) of an effective quality assurance performance assessment system for Agriculture in Botswana senior secondary schools given the available resources and constraints.

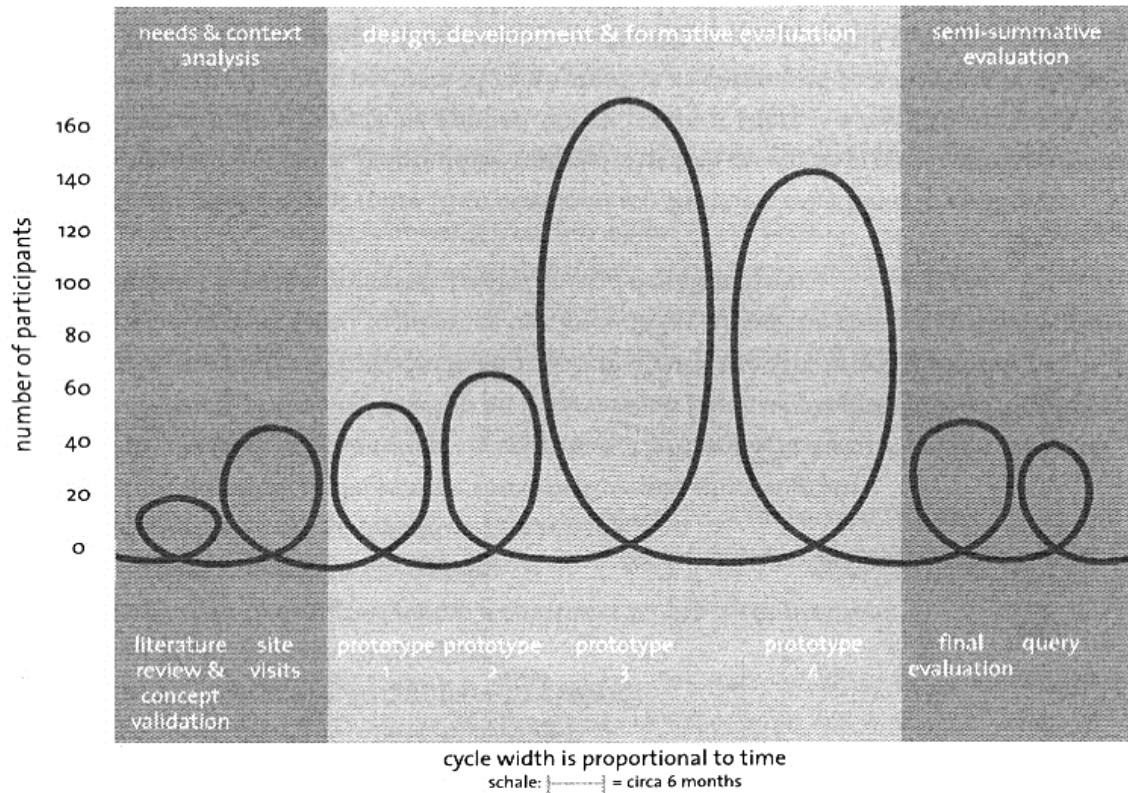
##### ***4.5.1 The nature of design-based research***

Design-based research, unlike conventional research designs, is a flexible methodology (Wang & Hannafin, 2005), which was aimed at improving educational practices through iterative analysis, design, development, and implementation that results in contextual-

sensitive design principles and theories (Brown, 2002; Collins, 2002). Throughout the design stage of the study, collaboration between researchers and practitioners in real-world settings was undertaken to develop ‘what works’ to solve complex problems in educational practice, and design principles that characterise the intervention (Cobb et al., 2003; Collins, Joseph & Bielaczyc, 2004; Gravemeijer & Cobb, 2006; Plomp & Nieveen, 2007).

The application of DBR in this study was appropriate to go beyond narrowly measuring students’ learning through paper-and-pencil tests, which essentially measures only one variable (Brown, 1999; Collins, 1999; Collins et al., 2004; Jan van der Akker, Gravemeijer, McKenney, Nieveen, 2006; Langmann & Shulman, 1999; Levin & O’Donnell, 1999). DBR tries to optimise how the interactions of different variables in a natural setting affect learning (Barab & Squire, 2004; Collins et al., 2004). This has the added burden of producing too much data, arising from the need to combine qualitative and quantitative methods, even though ultimately better understanding of learning is achieved (Brown, 1999; Collins, 1999; Collins et al., 2004).

The application of design research in this study drew heavily on the robust paradigms of pragmatism and to a lesser extent from that of constructivism (Walker, 2006), to help understand the relationships between educational theory, designed artefact, and practice (DBRC, 2003). The study remained flexible throughout to accommodate the ever-changing nature of natural settings (Mertens, 2010; Ornstein & Hunkins, 1993). Consequently, the development of the intervention was interactive and responsive to iterative stages of formative evaluation and re-designs (Bannon-Ritland, 2003), which involved multiple design-test-revise cycles as illustrated in McKenny’s (2001) CASCADE-SEA study presented in Figure 4.3.



*Figure 4.3: The cyclic process of design based research of the CASCADE-SEA study (source: Plomp & Nieveen, 2007. p.14 adapted from McKenny, 2001)*

Formative evaluation data helped to refine the initial prototypes and, in turn, to develop a more detailed design intervention (Collins et al., 2004) which applied to other settings for generating design knowledge or principles grounded in broader contexts. Formative evaluation thus has served different functions in the various stages of development and was built within each criteria of quality intervention. Formative evaluation can be perceived as having various layers in a design research study, as illustrated in Figure 4.4 (below). The evaluation comprised four layers, increasing in complexity from bottom to top.

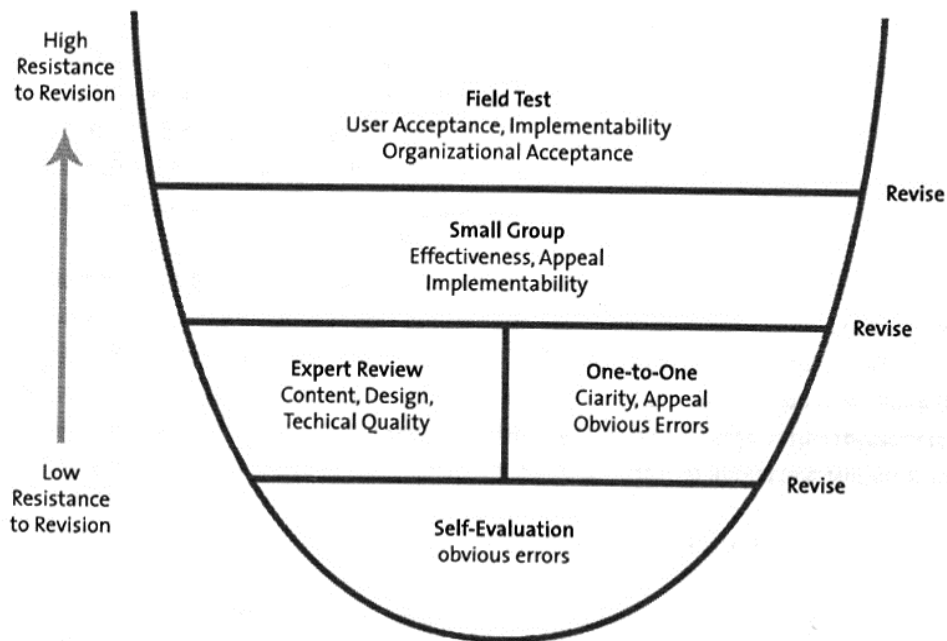


Figure 4.4: Tessmer's layers of formative evaluation (Source: Plomp & Nieveen, 2007.p.28 adapted from Tessmer)

The first layer was the evaluation of the developed prototype to check for obvious errors. The second layer comprised parallel evaluations made by experts for content, design and technical adequacy. This was carried out through interviews on a one-to-one basis, so as to clarify issues. During try-out, evaluation was made through small groups to gauge implementation and effectiveness of the intervention. The top of the diagram illustrates the high resistance during field test implementation, user acceptability and organisation acceptance. The high resistance at this level (top) is due to practitioners resisting change and preferring to work with the tried, tested and proven methods.

Successive refinement cycles resulted in the researcher revealing what did work, and how it worked under certain conditions in a specific context to generate well-supported design theories about learning and instruction (Collins et al., 2004). This reinforced deeper understanding of complex learning environments (Cobb et al., 2003). As Cobb et al. noted multiple sources of data from observations, interviews, surveys and documentations resulted in rigorous, empirically grounded claims and assertions (2003).

To produce high quality intervention, attention was paid to relevance (content validity), consistency (construct validity), practicality and effectiveness at different stages of the design

(Nieveen, 1999; Plomp & Nieveen, 2007). Table 4.3 succinctly summarises the evaluation criteria of quality intervention design.

*Table 4.3: Criteria for high quality intervention*

Criterion	Description of activities
Validity (also referred to as content validity)	There is need for intervention and its design to be based on state-of-the art (scientific) knowledge.
Consistency (also referred to as construct validity)	The intervention is ‘logically’ designed
Practicality	<b>Expected</b> The intervention is expected to be usable in the settings for which it has been designed and developed.
	<b>Actual</b> The intervention is usable in the settings for which it has been designed and developed.
Effectiveness	<b>Expected</b> Using the intervention is expected to result in desired outcomes
	<b>Actual</b> Using the intervention results in desired outcomes

*(source: Adapted from Plomp & Nieveen, 2007. p.94)*

The criteria of validity, for example, was achieved by ensuring that the components of the intervention were based on state-of-the-art knowledge and consistently linking all components to each other (Plomp & Nieveen, 2007) so that content validity could be high for all users irrespective of the different domains they refer to (McIntire & Miller, 2007). This criterion was emphasised during the needs analysis stage of the study (baseline survey), while less attention was given to practicality and effectiveness, which were emphasised at later stages of development.

#### **4.5.2 Research design**

Since this phase of the study aimed at improving the performance assessment programme of Agriculture in senior secondary schools in Botswana, design-based research was appropriate in designing an intervention to improve practice (Collins, Joseph & Bielaczyc, 2004; Gravemeijer 1998; Hoadley, 2002; The Design-Based Collective, 2003). The model was informed largely by Mafumiko's (2006), as depicted in Figure 4.2 (above). The development of the intervention was iterative, adopting a cyclic approach of design, evaluation and revision (Plomp 2008; Van den Akker, Branch, Gustafson, Nieveen & Plomp, 1999) and resulting in four prototypes. Developed in collaboration with practitioners at various stages of the design process, practitioners ensured that the intervention addressed practice and its success was measured by its practicality (utility) in real contexts (Gravemeijer, 2006). Subsequent prototyping ultimately contributed to substantive theory development (Barab & Squire, 2004; National Research Council, 2002; Plomp, 2008; Van der Akker et al., 2006).

#### **4.5.3 The research process**

Three performance tasks drawn from the BGCSE Agriculture curriculum were developed from the topic *Field Crop production*. Developing standard task and assessment materials from this topic was made to coincide with the schools implementation programme for minimal intrusion. Task 1 was based on *Preparing a plot and planting*. Task 2 on *Fertilizer Application as basal dressing*, and Task 3 on *Controlling weeds using chemicals*. The criteria for choosing tasks were discussed in Section 3.3.

The first prototype was developed and formatively reviewed by experts, mainly to validate content, and to some extent practicality (Plomp & Nieveen, 2007), as per criteria for high quality intervention outlined in Table 4.3. Relevant authorities on the topic were drawn from disciplines of Agriculture and Assessment, and their feedback was incorporated into the redesign of the second prototype, which was piloted in one government school. Each of the three tasks of the second prototype was administered to one class for practicality and feedback incorporated in the cyclic redesign of the third prototype.

The resulting third prototype, after incorporating feedback from teachers and students, was tried out in three different schools with each of tasks 1, 2 and 3 being implemented simultaneously in the same school, as shown in Table 4.4 (below). Three schools were the maximum possible number that could be used for try-out, to allow the researcher to move

around to observe assessment being conducted concurrently. For example, in school 1, one teacher implemented task 1, while another implemented task 2 and the third one implemented task 3 simultaneously.

*Table 4.4: Schedule of task implementation in schools*

School	Number of teachers implementing task:			Total of tasks in each school
	Preparing a plot	Fertiliser Application	Weed Control	
1	1	1	1	3
2	1	1	1	3
3	1	1	1	3

Formative evaluation to identify factors that prevented the intervention from meeting its stated targets was enacted by both teachers and their students. The aim was to evaluate the effectiveness and to a lesser extent the way the intervention operated in practice (Collins et al., 2004). The intervention was considered practical if users found it applicable with ease and compatible with the environment in which it was implemented (Persse, 2006), as well as with the developer’s intention (Plomp & Nieveen, 2007). The criterion of effectiveness was emphasised and measured by increased realisation of the desired outcome (Plomp & Nieveen, 2007).

The fourth prototype was designed and developed based on the outcome of the evaluation of the third prototype. However, this was to be reviewed in a workshop by the practitioners and evaluated for its effectiveness, resulting in the design and development of the final prototype, ready for enactment in the real field. This did not take place, as mentioned in Section 4.3. A more detailed discussion of each stage of development of the prototype is discussed in Chapters Six and Seven.

#### **4.5.4 Data collection**

Data collection was triangulated using different methods, namely observation, questionnaire, and interview, to obtain information from different sources such as students, teachers, records, experts and school administrators. Triangulation helped to check for consistency of evidence (Mertens, 2010).



**Observation:** both quantitative and qualitative observations were the most important assessment tool (Mindes, 2007) to be used during data collection. Observation was directed towards, inter alia, participants' social interaction, behaviour, informal interactions and unplanned activities, formal and planned activities, unobtrusive measures, and what does not happen (Cormack, 1991; Mertens, 2010). The instrument was made up of three sections (see Appendix 4.5), consisting of rating scales for quantitative observation (Bordens & Abbott, 2005; Goodman & Carey, 2004), namely *Instructional behaviour* with 15 items; *teacher's knowledge of assessment* with six items; and *availability of resources* with five items. The remaining aspect of the questionnaire consisted of items for qualitative observation. This enabled full description of behaviour (Ary et al., 2006; Dyer, 1995) to provide highly accurate, detailed, and verifiable information, not only about the person being assessed, but also about the surrounding context.

**Interview:** it was administered to gather data of respondents' opinions, feelings and beliefs (Ary et al., 2006; Forrester, 2010) and knowledge, values, experiences, ways of seeing and thinking and acting about the situation (Schostak, 2006) in their own words. During the interview, the researcher masqueraded as a participant-as-observer (Mertens, 2010), as the role was more peripheral during the process of data collection. Interview data were used to supplement the responses of the questionnaires (Alreck & Settle, 1995; Ary et al., 2006; Colton & Covert, 2007) and to verify information obtained through observation. Though personal interview is the most expensive type of interview, it was inevitable in this study because of the body language and the interviewees' contextual surroundings (Mertens, 2010), which were crucial for observation. The interview was semi-structured and conducted in a conversational style allowing for easy probing for understanding and additional information (Forrester, 2010; Mertens, 2010; Weisberg, Krosnick, & Bowen, 1996).

A **Focus group** comprising six students was formed and interviewed in each school where the intervention was enacted. The number for the focus group was kept low to ensure that all participants remained actively involved in the group discussion throughout the data collection phase (Willig, 2001). Members were selected with the help of the teachers, who identified individuals who were able to provide maximum insight and understanding of performance assessment (Ary, Jacobs, Razavieh & Sorensen, 2006). The interview was a semi-structured conversational type (Coolican, 2006), consisting of open-ended questions, which were posed as a guide to individual respondents (see Appendix 4.6). Responses were recorded using an

audio tape recorder to capture every participant's point (Fink, 2005). The student who was ready to answer raised his/her hand and continued to talk. This was meant to avoid chorus responses, which would make the transcription process difficult.

Like the students interview discussed above, the **teacher interview** was a semi-structured conversational type of interview (Coolican, 2006), consisting of open-ended questions, which were posed as a guide to individual respondents (see Appendix 4.7). The responses were recorded using an audio tape recorder. Interview data<sup>8</sup> was used to supplement the responses of the questionnaires (Alreck & Settle, 1995; Ary et al., 2006; Colton & Covert, 2007) and to verify information obtained through observation.

The **Student questionnaire** comprised seven questions in all. One question, with five sub-questions, was about student's demographic information such as sex, age, class, and school of the respondent (See Appendix 4.8). Another question was a Likert scale with 11 questions seeking students' opinions about the 'new way' of doing practicals. The rest of the questions were open-ended to give students the opportunity to suggest ways of improving the intervention design.

The **Teacher Questionnaire** consisted of eight demographic questions, with variables such as age, sex, teaching experience, qualification, class taught, and post of responsibility (See Appendix 4.9). It also consisted of five subscales of instructional behaviour, knowledge of assessment, standardising assessment, class management, and student attitudes. The subscales sought to measure the relative impact of the intervention as perceived by the participants. The questionnaire also consisted of open-ended questions to obtain in-depth information (Colton & Covert, 2007) about overall quality, content, format and language of the tasks.

**Document and content analysis** was made of students' detailed records of their activities during the practicals. Bogdan and Biklen, (2003) suggest that subject-produced data should be employed as part of studies where the major thrust is participant observation or interviewing. Records kept by students allowed the researcher access to information that would otherwise be unavailable (Fink, 2005; Mertens, 2010). To guide students on keeping records, a record book was produced with such contents as date, activity, tools/materials, and reasons for carrying out the activity (See Appendix 4.10).

---

<sup>8</sup> Whilst 'data' is the Latin plural of datum, it is often regarded as uncountable and may therefore be treated as singular for grammatical purposes, as in this study.

#### 4.5.5 *Data Analysis*

Quantitative data was analysed descriptively in terms of frequencies of each item response, percentages, means, and standard deviations to describe the distribution of scores. Data from open-ended items was analysed qualitatively, using thick descriptions to capture respondents' views and organizing them into themes.

All interviews were audiotaped and transcription done as a pre-requisite to data analysis by a professional (Cormack, 1991). Although transcriptions involved a considerable investment of time, it was inevitable as a step to organise data into themes. Units of information were identified to serve as the basis for defining categories (De Vos, 1998). Data was then reduced (coded) and sorted into specific categories (Wiersma & Jurs, 2005). Coding was done to accurately capture the information in the data relative to what was being coded, and to describe and understand the phenomenon being studied. Units applicable to each category were then compared and constant comparison of the units generated theoretical properties of the category (De Vos, 1998).

In-depth analysis was varied through participants' perceptions, understandings of meaning and interpretations (McDavid & Hawthorn, 2006). Thematic analysis identified words or phrases that summarized ideas conveyed in interviews or statements in a narrative (McDavid & Hawthorn, 2006) to produce substantive theory that pragmatically applied to the context (Forrester, 2010; Mertens, 2010).

Qualitative data analysis began soon after observation data collection commenced (Wiersma & Jurs, 2005), thus data collection and analysis were intertwined during observation. All interviews were transcribed in full and an iterative process of qualitative analysis was employed, drawing upon elements of grounded theory. Analysis began with open coding, assigning codes using respondent's own words as far as possible. A set of categories were then constructed which were thought to best describe interviewees' conceptions.

#### 4.6 METHODOLOGICAL NORMS

A number of strategies were employed to ensure that the inferences made from the study were valid and reliable. These strategies were aimed at eliminating extraneous variables to

make the results valid and reliable (Bordens & Abbott, 2005; Lincoln & Guba, 1999). The dependability of results is discussed in Subsection 4.6.1, while the rigour of research ethics is discussed in Subsection 4.6.2.

#### ***4.6.1 Dependability of the Results***

Studies in social research are not as reliable as in the physical sciences (Colton & Covert, 2007), hence their validity and reliability are somewhat different (see Section 3.7). Validity and reliability of the outcomes in this study were ensured by triangulating data collection strategies and sources (McDavid & Hawthorn, 2006).

Practitioners' and stakeholders' participation at various stages of standard tasks and assessment materials' development was one strategy used to strengthen dependability of results. Participants had a chance to identify and define the problem, resulting in equivalent tasks being developed, which measured a given content domain of importance to assess students (Mehrens & Lehman, 1991), despite their different culture and instructional context (Miller-Jones, 1989). Similarly, stakeholders' involvement in the development of the performance assessment materials increased significantly the probability of construct representation and lowering construct-irrelevant variance (Ary et al., 2006).

More assessment tasks were used through a matrix sampling design. Three different performance assessment tasks, which were equivalent, were administered to separate samples of students to improve reliability. Students work was scored by their teachers, who used clearly developed criteria to ensure that every student, irrespective of the geographical location, was scored the same (Airasian, 2008). These criteria were iteratively developed and formatively evaluated by teachers and other experts in assessment and in Agriculture.

Systematic and repeated observations carried out over varying conditions also increased the reliability of assessment (Mertens, 2010), as participants were asked to verify the information (member checking) before analysis and report writing commenced (Bogdan & Biklen, 2003). This ensured that the information they provided had been captured and edited accordingly. Furthermore, an analytic inductive methodology in observation to test emergent propositions (Alder & Alder, 1994) was followed, which had the effect of increasing reliability. Presentation of observational findings were written in such a way that the accounts contained a high degree of internal coherence, plausibility and correspondence to what readers

recognise from their own experiences and from other realistic and factual texts (Alder & Alder, 1994).

The researcher, as the main qualitative data collection instrument, was sensitive, adaptable and responsive to changing circumstances, and observed activities silently (Patton, 1990) so as not to influence the outcome.

#### **4.6.2 Ethical considerations**

Following research ethics is an important step in validating the outcomes. An application for ethics approval was submitted to University of Pretoria (UP) Ethics committee, detailing how the study would be conducted without violation of the rights and privacy of the participants. Upon approval by the UP Ethics Committee (see appendix 4.11), permission to conduct research in Botswana schools was sought from the Ministry of Education's Department of Planning, Research and Statistics (DPRS) and granted (see Appendix 4.12) for a year, and subsequently renewed after unsuccessful completion in the stipulated period. And clearance certificate was granted by UP for abiding by the code of conduct (Appendix 7.3).

The study's participants were both professionals (school administrators and teachers) and students in schools. Consequently, permission letters were written to Regional Education Officers seeking permission to conduct research in their schools. Permission was granted (see Appendix 4.13) paving way for seeking permission from schools (Appendix 4.14) and finally from individual participants (Appendix 4.15). Since some of the students were below the age of eighteen, permission was therefore sought from their parents or legal guardians for their consent to participate (Appendix 4.16). The students above eighteen years were requested to participate on a voluntary basis after thorough explanation of the study's objectives and likely benefits.

Professional participants were fellow colleagues. For them to make an informed decision of whether or not to participate, thorough explanation of the purpose of the study, benefits of participation, and potential risks or harm associated with participation in the study was undertaken. Confidentiality of their participation and protection of the information they volunteered to the researcher were guaranteed. Participants were also assured of the non-existence of the possibility of linking them to the information they provided, through the use of pseudonyms for follow-up and reporting purposes. They were informed that information

gathered would be used solely for the purposes of improving performance assessment and that no evaluation of their professional undertakings would be made or discussed.

In addition, participants were informed of their right to withdraw from the study without explanation or justification if they wished. Contact details of both parties were exchanged so that those interested in the report could access it, and they were informed that the researcher was available for further consultations. Upon completion of the study, the researcher declared to the Ethics Committee that the stipulated conditions had been abided by, and how the research data and/or documents were stored, resulting in the issuance of clearance certificate.

#### 4.7 CONCLUSION

This study employed a combination of descriptive research and design research considered appropriate since the aim was to understand and explore the characteristics and quality processes needed in the performance assessment of Agriculture Form Four students to ensure valid and reliable examinations in Botswana. The study was anchored in pragmatic and constructivist perspectives in which knowledge was ascertained by means of reason or experience in a constantly changing environment, as knowledge should be created in context, and applicable to the people concerned. Pragmatism and constructivism are therefore associated with learning advanced knowledge and skills in complex, ill-structured domains, whereby behaviour cannot be predicted, nor acceptable performance be precisely defined.

A baseline study preceded the development of the intervention, to identify the problem's context. This was important in understanding and describing assessment practices and processes, as well as points of views and attitudes that were held by practitioners. Findings of the baseline survey informed the iterative development of the intervention, carried out in collaboration with stakeholders at different stages, resulting in successive better prototypes due to incorporation of feedback from formative evaluation. Officer experts were involved during the evaluation of consistency, while teachers were involved at all the evaluation stages, and students were involved during the evaluation of practicality and effectiveness (see Table 4.3). The summative evaluation of the last version of the prototype was not done, as explained in Section 4.3.

## CHAPTER FIVE

### AGRICULTURE PERFORMANCE ASSESSMENT PRACTICES IN BOTSWANA

#### 5.1 INTRODUCTION

This chapter discusses the findings of the baseline survey conducted to understand the processes of performance assessment for certification. Specifically it addresses the first research question which sought to find out how valid and reliable is the performance assessment processes in Botswana. To understand the validity and reliability of the processes of performance assessment, Sub-questions (a) through (c) (See Section 1.6 and 4.1) guided this phase of the study.

The outcomes of the baseline survey underpinned the development of an intervention infusing quality assurance processes. Section 5.2 outlines the biographical data of the respondents which included age, sex, qualification, training, teaching experience and class size. Section 5.3 presents Agriculture performance assessment practices of teachers in Botswana schools such as mode of assessment; learning autonomy; assessment for learning; resources availability; standardisation of marking; supervision and monitoring of assessment; and attitude towards performance assessment.

Section 5.4 presents the discussion of findings of the study and Section 5.5 is the conclusion leading to Section 5.6, which examines the implications of the findings for the development of the intervention. The results presented in this chapter are mainly derived from survey questionnaires and supplemented by interviews.

#### 5.2 BIOGRAPHICAL DATA

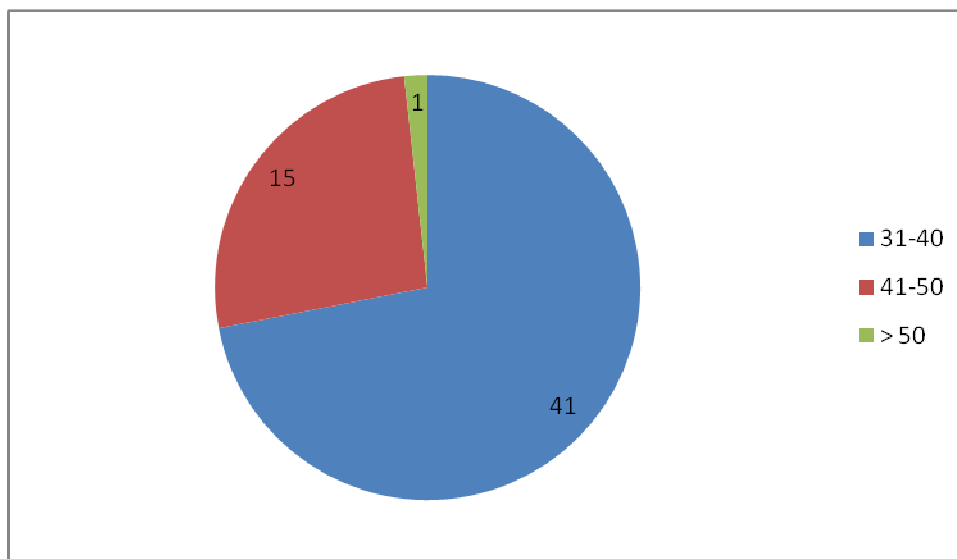
This section presents the biographical findings of the respondents. The response rate is as shown in Table 5.1 (below). Throughout the discussion of this study, senior teachers were treated as teachers unless specified.

*Table 5.1: The response rate of respondents*

Respondents	Expected	Attained
Teachers	68	57
Senior Teachers	13	11
Administrators	26	21

### **5.2.1 Teachers' age and gender**

All except one teacher who returned their questionnaires belonged to the active group of 31-50 years (combining categories of 31-40 and 41-50) as shown in Figure 5.1 (below). None was below the age of 31 years. There were 37 male teachers.



*Figure 5.1: Distribution of teachers' age (n = 57)*

### **5.2.2 Teachers' and school administrators' experience**

International research has shown that teacher experience in teaching their subject is one important factor for effective assessment (Broadfoot, 1994; Maxwell, 2004). Analysis showed that teachers were well experienced as almost all teachers (56 out of 57) had more than 5 years of teaching experience, as shown in Figure 5.2 (below). More than 5 years is considered adequate teaching experience by the system, because that was the minimum



experience for one to be considered for a post of responsibility. Senior teachers too were experienced as none had less than 5 years in their current post. On the other hand, all school administrators had at least 11 years teaching experience but none had more than 21 years of experience.

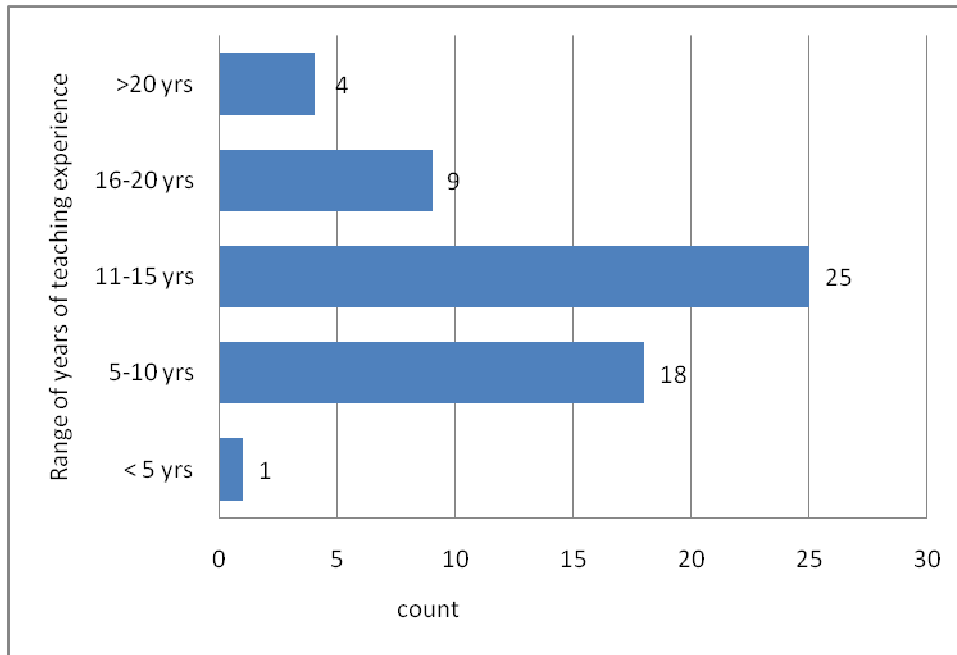
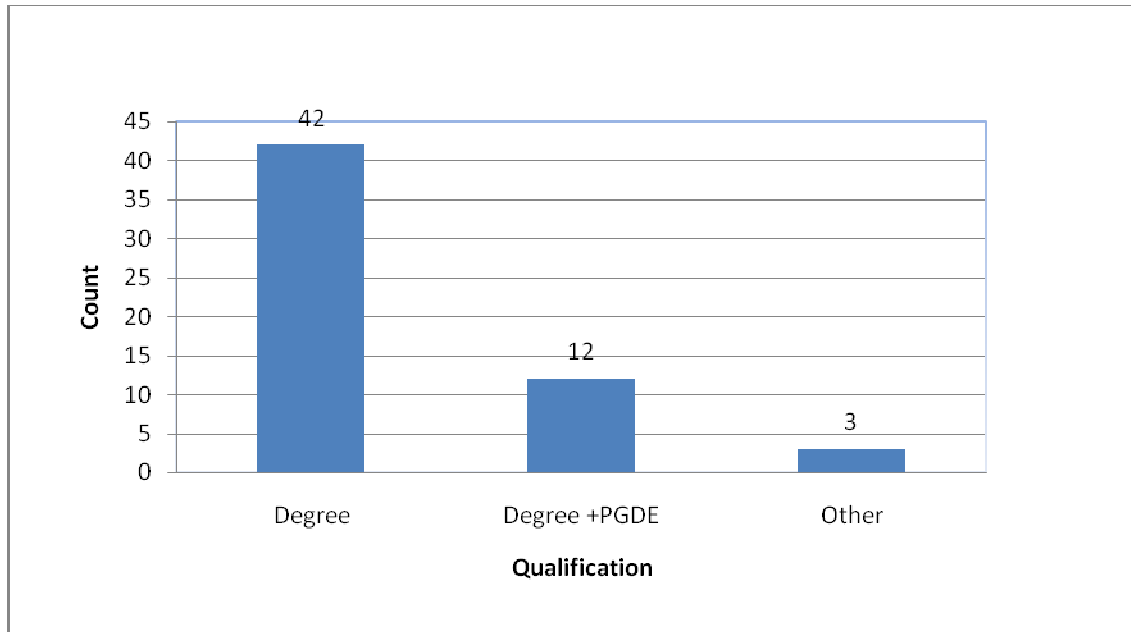


Figure 5.2: Teachers' teaching experience ( $n = 57$ ).

### 5.2.3 Teachers' and school administrators' qualification and training

The quality of teaching is heavily dependent on good quality training (Chong, 2009) (See Sections 2.10 and 3.5). Figure 5.3 shows teachers' qualification. The study revealed that all teachers had a degree qualification) which was a requirement to teach in a senior school (See Section 2.9). More than one-third had at least masters' qualification. School administrators were found to be equally well-qualified. Whereas only one school head had diploma qualification, the rest had at least a degree. However, it should be noted that qualification to teach is not in itself a sufficient condition for effective assessment, but rather training to assess is necessary to equip teachers with the necessary skills (Tindal & Haladyna, 2002).



*Figure 5.3: Teachers' qualifications (n = 57)*

The number of teachers and school administrators trained to conduct performance assessment are presented in Table 5.2. The results indicate that teachers and school administrators lacked training to conduct performance assessment. Only about one-third of teachers took a course in performance assessment during their training. About one-fifth to half of teachers had orientation on how to conduct performance assessment. Teachers who did a course on practical assessment are 19 while administrators are 9. Similarly teachers who did a course in assessment are 25 and 9 school administrators did the course. This is in agreement with Stiggins (2002) who noted that assessment in America is not considered a requirement to teach. Given this situation, Pellegrino, Chudowsky and Glaser, (2001) position that performance assessment course should be made compulsory to all student-teachers is welcome.

Table 5.2: Proportion of teachers and school administrators trained to conduct performance assessment

Statement	Teachers (n = 57)	School Administrators (n= 21)
a) I was inducted on how to conduct practical assessment when I started teaching.*	27	
b) I attended an in-service training sometimes in the past, on how to conduct practical assessment.	19	5
c) I did a course in assessment at College or University.	25	9
d) I did a course in performance assessment at College or University.	19	9
e) I was trained on how to develop performance tasks.*	11	
f) I was trained on how to develop scoring criteria/markings guide for scoring performance tasks.*	13	
g) I trained on how to use scoring criteria/markings guide when marking performance tasks.*	26	

\*Applicable to teachers only

Of particular concern is that cross-tabulation reveals that 25 teachers neither did a course on practical assessment nor a course on assessment during pre-service training as shown in Table 5.3. The current state of affairs is not good for the education system since teachers are the appropriate assessors of what is inaccessible to the external examination (Pellegrino, Chudowsky & Glaser, 2001). Teachers lacking training to assess can not be expected to effectively assess, if ever they do they are bound to concentrate on trivial outcomes (Tindal & Haladyna, 2002). However, teachers' technical competence to assess invariably facilitates the interpretation of performance criteria. Lack of skills and knowledge in assessment imply that teachers cannot develop appropriate materials for assessment consistent with the national curriculum as asserted by Kanjee and Sayed (2008).

The findings do not reflect a good picture on the implementation of RNPE which is the driving force of the education system in Botswana for a term of twenty-five years. In particular, recommendation 42 (b) of RNPE which calls for adequate training of teachers to

handle CA has not been fully implemented, confirming (Stiggins, 1997) assertions that teachers too who are in the forefront of assessment do not understand it. However, this does not imply that teachers cannot design and develop sound assessment given proper training and support resources.

Table 5.3: *Teachers who neither received training in performance assessment nor related training in assessment (n = 57)*

<b>Related training in assessment</b>	<b>Number of teachers</b>
Not inducted on practical assessment at university	16
Not attended in-service training on how to conduct practical assessment	17
Not done a course on assessment at college/university	25
Not trained on how to develop practical tasks	29
Not trained to develop scoring criteria	28
Not trained to use scoring criteria	20

#### **5.2.4 Class size**

The number of students in class has a bearing on the workload (Angrist & Lavy, 1999; Howie, 2006; Knostantopolous, 2008, Knostantopolous & Chung 2009). Table 5.4 (below) shows the frequency of students in a class. It was discovered that teachers taught between 2 to 6 classes. Senior Teachers taught fewer classes as they had administrative duties to perform in addition. The highest number of classes taught (6) translated to 24 periods per week in a 6-day-timetable, which was less than policy recommendation (6 x 4 periods - see Section 2.7 for elaboration) A sizeable proportion of classes (128) had a large number of students (41-50) far exceeding policy recommendation of 35 students per class. Jones (2006) commented that for effective instruction and assessment of performance tasks to yield better results (Finn et al., 2003), students should not exceed 25 in a class.

*Table 5.4 Frequency of Form Four Agriculture class sizes taught by respondents*

Class size	Frequency
20 or less	6
21 - 30	49
31 - 40	128
41 - 50	34

Class sizes should be reduced to manageable levels, since there is surplus of teachers in all subjects (Bennel & Molwane, 2008). Reducing class sizes would give teachers more time for individualised instruction and systematic observation to identify each student's needs and devise appropriate corrective actions instantly. As discussed in Section 3.4, low class sizes were a phenomenon of developed countries (Miller, Sen, & Malley, 2007).

### 5.3 PERFORMANCE ASSESSMENT PRACTICES OF TEACHERS

This section presents assessment practices of agriculture in senior secondary schools as perceived by both teachers and school administrators. The findings of the practices were important because they formed the basis for intervention development.

#### *5.3.1 The mode of assessment*

It was noted in Subsection 3.3.3 that different skills require different methods of assessment, and a mismatch occurs when a wrong method is used to assess a skill (Stiggins, 1997), resulting in inappropriately measuring students' achievement. Teachers were requested to rate themselves on a 5-point summative response scale, with 14 items, ranging from *Never (1)* to five *Always (5)*, regarding the appropriateness of methods they use to assess performance skills. Preliminary analysis prior to running principal component analysis using SPSS revealed good internal consistency of the scale, with Cronbach's alpha coefficient of .86 (Pearson, 2010). The item-total correlations ranged from .31 to .67, exceeding the minimum standard of .30 (Pearson, 2010). The highest correlation between two variables was 0.80, with the determinant of .001 surpassing the .00001 cut-off, indicating that variables correlated fairly well with each other. Thus there was no singularity or multicollinearity

among the variables. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .72 exceeding the cut-off suggested by Hutchinson and Sofroniou (1999) and Hair, Anderson, Tatham, and Black, (1995). The Bartlett's test of sphericity was significant ( $p < .05$ ), implying that the correlations for the data were adequate for factor analysis to yield distinct and reliable results (Meyers, Gamst & Guarin, 2006).

Items' means, standard deviations, factor correlations, communality estimates, and item-total correlations are presented in Table 5.5 (below). The means ranged from  $M = 4.18$  ( $SD = 1.24$ ) to  $M = 1.68$  ( $SD = 1.21$ ). Those with lower means (below 2.5) indicate that the practices rarely take place. Based on this, four scenarios emerged; (i) desirable practices which always occurred, (ii) desirable practices which rarely occurred, (iii) undesirable practices which rarely occurred, and (iv) undesirable practices which always occurred. Scenarios (i) and (iii) are practices characterising performance assessment and scenarios (ii) and (iv) should be eliminated from performance assessment practices. Fortunately, they were a few of latter scenarios.

Factor analysis produced communalities which were fairly high for each of the 14 items, ranging from .44 to .82, indicating substantial contribution to the component/factor solution. All variables had factor loadings of at least .40. Using the Kaiser-Guttman retention eigenvalues of greater or equal to 1.0, a three-factor solution provided the closest extraction. These three factors accounted for about 60% of the total variance. Factor 1 had Cronbach's coefficient alpha of .83, depicting *holistic assessment* and accounted for 25%. Factor 2 had Cronbach's coefficient alpha of .82, depicting *Marginal assessment*, and accounted for 18%, while factor 3 had Cronbach's coefficient alpha of .67, depicting *Multiple-rating*, accounting for 15% of the variance. The number of items, eigenvalues and variance accounted for, for each of these factors are presented in Table 5.6.

Table 5.5: Summary of items and factor loadings from principal components analysis with varimax rotation for mode of assessment ( $n = 57$ )

Item Name	Mean	SD	Factor loading			Communality ( $h^2$ )	Corrected Item- Total Correlation
			1	2	3		
Reassessing the same skill when the student did not do well the first time.	3.89	1.15	.80			.68	.61
Assesses each student more than once on the same skill	2.19	1.30	.72			.53	.47
Gives the same score to everyone in the group.	3.11	1.71	.65			.44	.57
Assess students' written practical test.	2.44	1.30	.62			.50	.60
Assesses students' records of practical work.	4.18	1.24	.61			.45	.53
Assesses students group work during practicals.	3.30	1.31	.54			.52	.65
Assesses the students' affective skills towards practical work.	3.21	1.37	.53			.57	.67
Assesses all students in a class one day on the same skill.	2.93	1.35		.90		.82	.71
Assesses all students in a class on the same skill	3.67	1.43		.87		.77	.79
Assesses students when working on practicals	3.88	1.39		.62		.61	.63
Gives each student a different score in a group.	3.88	1.43			.76	.67	.31
More than one teacher assessing same practical skill.	2.84	1.54			.68	.69	.40
Assesses students work only when they have completed.	1.68	1.21			.59	.55	.45
Works with another teacher to assess the students.	1.86	1.16			.56	.62	.50

Table 5.6: Characteristics of factors for modes of assessment

Factor Name	No. of items	Coefficient Alpha	Eigenvalue	Variance Accounted for
Holistic assessment.	7	.83	3.57	25.52 %
Marginal	3	.82	2.64	18.85 %
Multiple-rating	4	.67	2.20	15.72 %

Further analysis using an independent-sample t-test compared the region mean scores for each of the three factors, and there was no significant difference between the two regions on the frequency of emphasising holistic assessment ( $t = .432$ ,  $df = 55$ ,  $p < .05$ ), marginal ( $t = -1.11$ ,  $df = 55$ ,  $p < .05$ ) and multiple-rating of individual student ( $t = -2.02$ ,  $df = 55$ ,  $p < .05$ ). Since the education system is controlled from the central point, resources are distributed equitably. Although schools report to their respective Regional Education Officers, regions report to the same Permanent Secretary of the MoE&SD. Therefore the schools are likely to be uniform in all respects. Given that all Agriculture teachers had at least a degree qualification (See Subsection 5.2.3), not much variation in their pedagogical practices was expected.

### 5.3.2 Learning autonomy

The application of constructivism instructional approaches facilitates students' engagement in the construction of meaning and learning through active involvement (See Section 4.2). To gauge the extent to which constructivism strategies were entrenched in teachers' classes, teachers were requested to rate themselves on a 5-point summative response scale, with eight items, ranging from *Never (1)* to *To a large extent (5)*. Preliminary analysis revealed that the instrument was internally consistent with Cronbach's coefficient alpha of 0.84, and the highest coefficient between two variables being 0.65 and the determinant of 0.036, suggesting no singularity or multicollinearity among variables. The KMO measure of sampling adequacy was .71, and the Bartlett's test of sphericity was significant ( $p < .05$ ). The corrected item-total correlation ranged from .35 to .64, exceeding the minimum standard of 0.3.

Table 5.7 (below) presents the items' means, standard deviations, factor loadings, communality estimates, and item-total correlations for the scale. The means ranged from  $M =$



1.53, (SD = 1.07) to  $M = 4.11$ , (SD = 1.45). Only four instructional activities had a mean higher than the average of 2.50. This suggested that generally learning autonomy was moderately practiced by teachers in performance assessment. Factor analysis revealed fairly high communalities for each of the 8 items, ranging from .50 to .73. All variables had factor loadings of at least .56. Using the Kaiser-Guttman retention eigenvalues of greater or equal to 1.0, a two-factor solution provided the closest extraction. These two factors accounted for 66% of the total variance.

The number of items, Cronbach's coefficient alpha, eigenvalues and variance accounted for, for each of these factors are presented in Table 5.8. Factor 1 had Cronbach's coefficient alpha of .81 depicting *Peer Assessment* and accounted for 36% of the variance, while factor 2 had Cronbach's coefficient alpha of .73, depicting *Involvement of students in decisions making*, and accounted for 24% of the variance.

Table 5.7: Summary of Items and factor loadings from principal components analysis with varimax rotation for learning autonomy (n=57)

Item Name	Mean	SD	Factor Loading		Communality ( $h^2$ )	Corrected Item- Total Correlation
			1	2		
I provide guidance to help students assess one another's practical work	1.86	1.30	.85		.74	.64
Students are given opportunities to assess one another's practical learning	1.79	1.32	.80		.66	.60
Students are given opportunities to decide how they will be assessed	1.53	1.07	.76		.62	.60
I provide guidance to help students assess their own practical work	2.74	1.59	.66		.50	.55
I give students feedback after assessing/marking their practicals	2.46	1.55		.81	.73	.81
Students come up with their topics of study for the project	2.91	1.47		.75	.56	.56
I give the students chance to discuss how they learn in practicals	3.44	1.68		.56	.55	.56
I agree with students to assess them in practicals when they are ready	4.11	1.45		.56	.51	.56

Table 5.8: Variance accounted for by the two-factor solution

Factor Name	No. of items	Cronbach's coefficient alpha	Eigenvalue	Variance Accounted for
Peer Assessment	4	.81	3.80	36.00 %
Involvement of students in decisions making	4	.73	1.07	24.76 %

An independent-samples t-test compared the mean scores for the regions and found no significant difference in the extent to which the regions *involve students in decision making* in assessment ( $t = -1.383$ ,  $df = 55$ ,  $p < .05$ ), and the extent of *Peer Assessment* ( $t = .306$ ,  $df = 55$ ,  $p < .05$ ). Similarly, no significant difference was observed between male and female teachers on the extent of both *Involving students in decision-making in Assessment* ( $t = .869$ ,  $df = 55$ ,  $p < .05$ ), and extent of *Peer assessment* ( $t = -.045$ ,  $df = 55$ ,  $p < .05$ )

### 5.3.3 Assessment for Learning

Assessment for learning is intended for the teacher to diagnose students' strengths and weaknesses and provide differential instructional strategies according to their needs (ARG, 2002). Assessment for learning involves identifying what students have reached in their learning, what skills and knowledge are being established, and what skills and knowledge are not yet within the zone of Proximal Development (Vygotsky in Eysenck, 2004). However, Black and Wiliam, (1998a, 1998b) and Izard (1998) contend that practical implementation of assessment for learning to improve teaching and learning has been inadequate.

To determine teachers' understanding of assessment for learning, they completed a 5-point summative response scale consisting of eleven items, ranging from extremely *unimportant* (1) to *extremely important* (5).. A score of 5 represented a strong understanding while a score of 1 indicated a weak understanding. One item was negatively worded and had to be reversed before analysis.

Preliminary analysis revealed that the instrument was internally consistent with Cronbach's coefficient alpha of .91. The highest coefficient between two variables was .82 and the determinant was .01, indicating no singularity or multicollinearity among variables. The KMO measure of sampling adequacy was .86, and the Bartlett's test of sphericity was

significant ( $p < .05$ ). Corrected item-total correlation ranged from .31 to .81, satisfying the minimum standard of .30.

Table 5.9 (below) presents the items, means, standard deviation, factor correlations, communality estimates, and item-total correlations of the scale. The mean scores ranged from  $M = 3.33$  ( $SD = 1.42$ ) to  $M = 4.12$  ( $SD = 1.30$ ). Teachers understood the importance of formative assessment with feedback to help students learn and improve learning, although it was found that they moderately practiced learning autonomy.

Factor analysis produced communalities which were fairly high for each of the 11 items, ranging from .50 to .80. All variables had factor loadings of at least .65. Using the Kaiser-Guttman retention eigenvalues of greater or equal to 1.0, a one-factor solution provided the closest extraction. This one factor solution was non-robust because it accounted for only 46% of the variance, which is less than 50% for a robust one. There were nine items with an eigenvalue of 5.99. Cronbach's coefficient alpha for this factor was .92, indicating good subscale reliability.

One-way between-subjects ANOVA compared teachers experiences (10 or less years, 11-15 years, and above 15 years) on the importance attached to *Assessment for Learning*. There was no significance difference found between the groups  $F(2, 53) = 1.51$ ,  $p < .05$ .

Table 5.9: Summary of items and factor loadings from principal components analysis with varimax rotation for assessment for learning (n=57)

Item name	Mean	SD	Factor loading		Communality ( $h^2$ )	Corrected Item- Total Correlation
			1	2		
Modifying my practice in practical assessment in light of evidence from self-evaluations of my classroom practices.	4.02	1.18	.88		.80	.82
Modifying my practice in practical assessment in light of feedback from my senior teacher or other colleagues.	3.33	1.56	.84		.71	.63
Discussing learning objectives for practicals with students in the way they understand.	3.47	1.56	.81		.67	.73
Viewing students' effort as important when assessing their practicals.	3.33	1.42	.75		.73	.80
Encouraging students to view mistakes as valuable learning opportunities.	4.04	1.10	.71		.69	.77
Modifying my practice in practical assessment in light of feedback from my students.	3.96	1.25	.70		.54	.66
Valuing students' errors for the insights they reveal about how they are thinking.	3.58	1.48	.69		.50	.61
The outcome of students' assessment of practical tasks consisting primarily of marks and grades.	4.09	1.29	.66		.54	.66
Helping students to understand the learning purpose of each practical lesson	3.60	1.35	.65		.75	.78
Helping students to find ways of addressing problems they have in their practicals.	4.12	1.30		.84	.74	.48
Identifying students' strengths and advise them on how to develop them further.	3.91	1.47		.82	.67	.35

#### 5.3.4 Availability of Resources

Student-centred learning discussed in subsection 5.3.2 is feasible where there are sufficient resources for both students and teachers to facilitate practical work by students. Resources necessary for performance assessment are physical resources (infrastructure and tools/equipment), human resources and time. Teachers were requested to indicate the availability of resources in their schools on a 5-point scale ranging from Strongly Disagree (1) to Strongly Agree (5). A score of 5 represented a strong endorsement while a score of 1 indicated a weak endorsement about the availability of resources. There were 12 items in all, of which 5 were negatively worded and were reversed before analysis.

Preliminary analysis revealed that the instrument was internally consistent with Cronbach's coefficient alpha of .61. The highest coefficient between two variables was .60 and the determinant was .030 suggesting no singularity or multicollinearity among variables. The KMO measure of sampling adequacy was .66, and the Bartlett's test of Sphericity was significant ( $p < .05$ ).

Table 5.10 (below) presents the items, means, standard deviation, factor correlations, communality estimates, and item-total correlations. The means ranged from  $M = 1.23$  ( $SD = .85$ ) to  $M = 4.40$  ( $SD = 1.27$ ). Teachers' indicated that resources were not adequate to facilitate the conduct of practicals in schools. They endorsed only one resource to be adequately available, namely garden space. Communalities were fairly high for each of the 12 items, ranging from .53 to .81. All variables had factor loadings of at least .55 and corrected item-total correlation ranged from .09 to .54. Using the Kaiser-Guttman retention eigenvalues of greater or equal to 1.0, a four-factor solution provided the closest extraction.

Table 5.10: Summary of items and factor loadings from principal components analysis with varimax rotation for availability of resources (n=57)

Item name	Mean	SD	Factor loading				Communality ( $h^2$ )	Corrected Item- Total Correlation
			1	2	3	4		
Agriculture practical marking is scheduled in the timetable, separate from teaching time.	1.26	.88	.84				.72	.37
The marking of students' projects by teachers is officially allocated time.	1.81	1.36	.77				.63	.34
Agriculture practicals are scheduled in the timetable, independent of teaching time.	1.23	.85	.75				.60	.26
The Agriculture curriculum is loaded with content.	4.40	1.27		.83			.70	.17
Student/teacher ratio for Agriculture is high.	4.39	1.37		.80			.66	.19
Technical staff should be hired to help teachers during practicals.	4.05	1.46		.66			.57	.14
Animal structures are enough for all students doing Agriculture.	1.68	1.14			.85		.81	.41
Equipments/tools are enough for all students during practical lessons.	2.12	1.26			.72		.53	.21
There are enough animals for practicals for all students.	1.51	.89			.55		.71	.54
Teachers' workload is high.	4.28	1.28				.73	.58	.10
Garden space is enough for all the students.	2.89	1.66				.73	.70	.28
There are too many practicals done in Agriculture.	3.19	1.45				.57	.53	.09

Table 5.11 (below) presents factors and their items, Cronbach’s coefficient alpha, eigenvalues and variance accounted for, for each of the factors. Cronbach’s alpha coefficient was low for factor 4, thus items were not consistent with each other. Factor 4 was dropped, resulting in a three-factor solution. These three factors accounted for 51% of the total variance. Factor 1 depicted *Time availability* and accounted for 20% of variance, factor 2 depicted *Workload* and accounted for 16% of variance, factor 3 depicted *material resources* and accounted for 14% of variance, while.

Table 5.11: *Characteristics of factors for availability of resources*

Factor Name	No. of items	Cronbach Coefficient alpha	Eigenvalue	Variance Accounted for
Time availability	3	.80	3.16	20.56 %
Workload	3	.69	2.10	16.65 %
Material resources	3	.75	1.46	14.51 %
Factor 4	3	.48	1.02	12.70 %

### 5.3.5 *Monitoring and Supervision*

Monitoring and supervision is important for adherence to standards and to act judiciously in instituting corrective actions, as dependence on psychometric properties alone to guide one to the standards is no longer considered satisfactory (Wild & Ramaswamy, 2008). Scores should be valid before subjecting them to various forms of moderation procedures. In schools, senior teachers’ are the first line of monitoring quality, as they report matters relating to non-compliance in assessment directly to school management through School Heads who, by virtue of their positions, are Chief Invigilators. School management therefore has an important role to play in ensuring quality in performance assessment (Mamary, 2007).

Senior Teachers were asked to rate the frequency of monitoring and supervision on a scale ranging from 1=*Never* to 5=*Always*, while school management was asked to rate its understanding on various issues related to performance assessment. Preliminary analysis resulted in removing three items from the Senior Teachers’ scale due to their low item-total correlation coefficients, resulting in the satisfactory Cronbach’s coefficient alpha of .84. The highest coefficient between two variables was .85, while the determinant was .000332, suggesting no singularity or multicollinearity among variables. The KMO measure of sampling



adequacy was .31, and the Bartlett's test of sphericity was significant ( $p < .05$ ), suggesting adequacy of factor analysis to proceed.

Table 5.12 (below) presents the items, means, standard deviations, factor correlations, communality estimates, and item-total correlations for the scale. The means ranged from  $M = 1.90$  ( $SD = .92$ ) to  $M = 4.20$  ( $SD = .79$ ). Although the results indicated that supervision by senior teachers was frequent, it was confined to paper-work with little physical visits to the field (garden). Communalities were fairly high for each of the 9 items, with a range of .66 to .93. All variables had factor loadings of at least .62, while the corrected item-total correlation ranged from .42 to .84.

Using the Kaiser-Guttman retention eigenvalues of greater or equal to 1.0, a one-factor solution provided the closest extraction. This one-factor solution, with five items and an eigenvalue of 4.20, accounted for 38% of the total variance. However, this was not sufficient to provide a robust solution (Tabachnick & Fidell, 2001), since it accounted for less than 50%. Cronbach's coefficient alpha for this factor was .87, indicating good subscale reliability. An independent-samples t-test compared the mean scores for the two regions and found no significance difference on the frequency of monitoring and supervision by senior teachers ( $t = -.428$ ,  $df = 8$ ,  $p < .05$ ).

Table 5.12: Summary of items and factor loadings from principal components analysis with varimax rotation for monitoring and supervision (n=57)

Item name	Mean	SD	Factor loading			Communality (h <sup>2</sup> )	Corrected Item- Total Correlation
			1	2	3		
I hold meetings to discuss problems teachers face in carrying out assessment.	3.20	1.03	.92			.90	.84
I appraise Agriculture teachers during the assessment of practicals.	2.20	.92	.87			.90	.74
Supervise teachers when conducting practical assessment.	1.90	.92	.84			.88	.70
I check progress on project write-up.	4.20	.79	.74			.69	.57
I observe teachers assessing student practicals.	2.50	1.18	.62			.66	.42
I demand practical assessment marks every term for safe keeping.	3.30	1.64		-.76		.93	.49
I hold meeting to discuss expectations regarding practical assessment.	3.10	1.20		-.69		.83	.55
I check progress on teachers' practical assessment.	3.10	.74			.705	.90	.53

### 5.3.6 Standardisation of marking

Among other things, standardisation involves preparing scoring rubrics in advance; specifying clearly what and how to be assessed; and training teachers in psychometrics to interpret the criteria properly. Standardisation is necessary to apply the assessment criteria in the same way from one situation to the other so as to achieve fairness in scoring. Teachers were requested to indicate on a 5-point summative response scale ranging from *Never* (1) to *Always* (5), the extent to which they conducted internal standardised testing, before scoring students work..

Prior to running analysis with SPSS, the data was screened through descriptive statistics, and analysis revealed that the scale was internally consistent with Cronbach's coefficient alpha of .83. The highest correlation coefficient between two variables was .74, and the determinant was .022, suggesting no singularity or multicollinearity among variables. The KMO measure of sampling adequacy was .70, indicating that the data was adequate for principal component analysis. Similarly, the Bartlett's test of sphericity was significant ( $p < .05$ ), indicating sufficient correlation between the variables to proceed with the analysis. The corrected item-total correlation ranged from .44 to .72.

Table 5.13 (below) presents the items, means, standard deviations, factor correlations, communality estimates, and item-total correlations. The means ranged from  $M = 1.44$  ( $SD = 1.15$ ) to  $M = 4.37$  ( $SD = 1.51$ ). The results suggested that there seemed to be high standardisation before marking, but what was deficient was the involvement of school administration in the process. Teachers too marked their own students' projects, something that could contribute to lowering the validity of scoring. Communalities were fairly high for each of the 8 items, with a range of .52 to .89, indicating that each variable contributed substantially to the component/factor solution.

All variables had factor loadings of at least .64, demonstrating high correlation with their factors. Using the Kaiser-Guttman retention eigenvalues of greater or equal to 1.0, a one-factor solution provided the closest extraction. However, this was not sufficient to provide a robust solution (Tabachnick & Fidell, 2001), since it accounted for less than 50%. This one-factor solution had six items with an eigenvalue of 3.60 accounted for 43% of the total variance (Tabachnick & Fidell, 2001). Cronbach's coefficient alpha for this factor was .85, indicating good subscale reliability.

An independent-samples t-test revealed no statistical significant difference between the teachers who have experience in moderation and those who did not have experience in moderation on the extent of *standardising marking* ( $t = .52$ ,  $df = 55$ ,  $p < .05$ ). Likewise, no significance difference was observed between teachers who had experience in marking final examinations and those who did not have experience in marking final examinations ( $t = .1.67$ ,  $df = 54$ ,  $p < .05$ ), on the extent of *standardising marking*. It seems teachers who were constantly engaged by the examining body to moderate and mark final examinations never transferred the skills they acquired to their work places. This goes to show the extent of secondary treatment accorded to performance assessment, and revealed that internal monitoring and supervision structures were not efficient in terms of performance assessment.

Table 5.13: *Summary of items and factor loadings from principal components analysis with varimax rotation for standardisation of marking*

(*n* = 57)

Item name	Mean	SD	Factor loading		Communality (h <sup>2</sup> )	Corrected Item- total Correlation
			1	2		
We use the marking criteria from the Ministry of Education when marking the project report.	4.46	1.40	.81		.70	.72
The senior teacher ensures marking is done according to standard.	3.77	1.62	.78		.60	.56
We meet to discuss project documents from the Ministry, e.g. marking guide.	3.91	1.58	.78		.62	.49
We standardize internal marking of practicals.	3.75	1.66	.78		.61	.60
We standardize marking of project report.	4.37	1.51	.77		.59	.57
We use the marking criteria from the Ministry of Education when marking the practicals.	4.21	1.54	.64		.52	.58
The Chief invigilator attends our standardization sessions.	1.74	2.21		.94	.89	.16
We swap classes for internal marking of the project report.	1.44	1.15		.90	.83	.44

### 5.3.7 *Attitude towards performance assessment*

By nature, Agriculture is a practical subject which requires one at some point to be working in strenuous and untidy conditions, which could give rise to negative attitudes in some. Teachers were requested to rate their students', fellow teachers', and administrators' attitudes towards performance assessment on a 5-point summative response scale ranging from *strongly disagree* (1) to *strongly agree* (5).

Preliminary analysis revealed internal consistency of the instrument with Cronbach's coefficient alpha of .56. The highest coefficient between two variables was .81 and the determinant was .035, suggesting no singularity or multicollinearity. The Kaiser-Meyer-Olkin (KMO) value was .49, and the Bartlett's test of sphericity was significant ( $p < .05$ ). Corrected item-total correlation ranged from .10 to .51, with only 5 items out of eleven having corrected item-total correlation coefficient of acceptable level of more than .30.

Table 5.14 (below) presents the items, means, standard deviations, factor correlations, communality estimates, and item-total correlations for factor analysis. The means ranged from  $M = 1.84$  ( $SD = 1.25$ ) to  $M = 3.47$  ( $SD = 1.51$ ). Communalities were fairly high for each of the 11 items, ranging from 0.28 to .86. All variables had factor loadings of at least .29. Using the Kaiser-Guttman retention eigenvalues of greater or equal to 1.0, a two-factor feeble solution provided the closest extraction.

Table 5.14: Summary of items and factor loadings from principal components analysis with Varimax rotation for perception towards performance assessment ( $n = 57$ )

Item name	Mean	SD	Factor loading				Communality ( $h^2$ )	Corrected Item- Total Correlation
			1	2	3	4		
Agriculture is considered to be for the less able students by the other teachers?.	3.32	1.66	.89				.86	.51
Agriculture is considered to be for the less able students by other students?.	3.18	1.56	.89				.82	.42
Agriculture is considered last during allocation of students by the curriculum committee?.	2.86	1.49	.61				.65	.33
Students have a positive attitude towards practical work.	2.58	1.38		.84			.74	.18
Students in this school enjoy learning Agriculture.	2.96	1.34		.82			.68	.38
Agriculture is allocated enough money for practicals by the Ministry.	2.09	1.26		.59			.59	.13
Teachers feel that practicals take too much of students' time.	2.47			.44			.28	.26
School administration and the rest of staff believe that all students are capable of doing practicals.	3.47	1.51			-.73		.62	-.23
Teachers feel that Agriculture should be taught theoretically only.	1.84	1.25			.61		.40	.10
Agriculture is treated as a non practical subject by school administration.	3.07	1.64				.85	.78	.18
Students refuse to do practical work.	2.88	1.30				.46	.57	.47

Table 5.15 (below) presents the factors and their items, eigenvalues, Cronbach's coefficient alpha and variance accounted for.

Table 5.15: *Characteristics of factors for perception towards performance assessment*

Factor Name	No. of items	Cronbach's coefficient alpha	Eigenvalue	Variance Accounted for
Filler Subject	3	.78	2.66	20.07 %
Positive attitude	4	.72	1.98	19.04 %
Factor 2	2	-	1.27	12.51 %
Factor 3	2	-	1.06	11.76 %

These two factors accounted for 39% of the total variance. Cronbach's coefficient alpha for factor 1 was .78 and for factor 2 was .72. Factor 1 had a Cronbach coefficient alpha of .78 and was named *Second-class assessment*, factor 2 had Cronbach's coefficient alpha of .72 was named *Positive attitude*.

#### 5.4 DISCUSSION

The main aim of conducting the survey, as discussed in Section 4.4, was to determine the validity and reliability of performance assessment processes in Botswana, by understanding how performance assessment is done in relation to the policy. Findings revealed that teachers were not well trained to handle performance assessment and majority of teachers did not even receive induction related to assessment. Although a few classes had more students than recommended by the policy, on the contrary, majority of classes had more students than the international average (Jones, 2006).

Because of insufficient training, assessment of performance tasks in schools was found to be inclined towards product assessment as echoed by one teacher: *Sometimes I give a general mark for the product, though I know it's wrong, but there is nothing I can do because the tool we use is not clear, and we have large class sizes and loaded curriculum.* Although assessment of the product is necessary and desirable, over-emphasis without understanding how learning took place is more a case of conducting an activity that merely audits learning



(Shepherd, 2000; Wiggins, 1993). Gronlund (2003) outlined situations when each of product and process assessment should be carried out (see Section 3.3).

Process assessment allows the students to demonstrate in a variety of ways their competence in using knowledge and skills learnt from different areas (Gronlund, 2003; Popham, 2005). This promotes improvements in learning and excellence as the ultimate goal of assessment (Wiggins, 1998). However, processes were not frequently assessed due to lack of training, limited time, insufficient resources, lack of standardised criteria, lack of support and high workload. Teachers' understood very well the importance of assessing processes as one teacher retorted: *We are expected to mark as they are working but unable to do so because we end up assessing product after lessons.*

Even though product assessment dominated the assessment process, it was not appropriately carried out. For example, there were no standard criteria that were used throughout the country. Each school devised its own assessment criteria based on the outline provided in the syllabus. Lack of standardised criteria for scoring implies that the assessment instantly becomes unreliable as teachers indicated: *Our assessment criteria needs to be standardised throughout the whole country so that when we say we give a certain mark for a skill, it should be the same, but I don't think that is the case now.*

Assessment was primarily done by one teacher despite well documented evidence of improved reliability when multiple raters are involved (Airasian & Russell, 2008; Rennert-Ariev, 2005). Multiple rating is desirable since a single assessment could not be relied upon for a variety of reasons such as illness, family problems, or other distractions. Performance at a single time may not be regarded as representative of the student capabilities. Even Testing Companies caution against making important decisions based on a single test score (McMillan, 2000).

Teachers assessed all students at the same time, without using clear criteria. This is uncharacteristic of performance assessment. In some cases, a group score was assigned whenever students were doing group work, and in most cases such marks were inflated with the aim to pass the students. Such assessment resulted in failure to elicit from the students the most advanced performance of which each was capable. As a consequence, assessment was not carried out to diagnose students' state of learning, but rather to satisfy the requirement of

the Awarding Body, as teachers felt it was imposed on them: *I wouldn't say marks are dependable; we are doing it as a requirement.*

Although assessment concerned students, they were not involved in assessment decision-making, and neither did they know in advance what or how they would be assessed. Such assessment was teacher-centred, with the teacher on one hand directing everything and students on the other receiving. Involving students in their own assessment allows them to know in advance what and how they would be assessed (Black & William, 1998), thus making assessment more realistic and educative (Wiggins, 1998). Harlen (2006) posit that if students know how assessment is done they can use the criteria to evaluate their own work prior to the teacher's evaluation and so improve their learning.

Performance assessment was given secondary treatment as many felt it should be done through paper and pencil. Students too had negative attitudes towards performance assessment. They felt it made them dirty and involved a lot of work, and suggested that people should be hired to do the practicals for them. This revealed their lack of understanding of the objective of performance assessment. However, not every student viewed performance assessment negatively. Some viewed it as developing their creative thinking and imparting life skills that would be useful after school life. As for the school management, it was said to be very supportive in trying to instil positive attitudes in students towards practicals. One teacher commented thus: *"they assist a lot; they also talk to the students if there is a problem, and even involve parents when they fail to resolve the issue"*.

As indicated earlier, senior teachers and school administrators' monitoring and supervision was not thorough, and teachers took advantage of the situation to award marks, even where they were not due. For example, one teacher said: *"Teacher's feel that this is the area they can influence the final grade of the students. They tend to increase the marks of the students. There is a lot of subjectivity"*. Such an act should be vehemently condemned as it degrades the professionalism of teachers and teaching as a profession. To illustrate the extent of insufficiency of school administration monitoring, one senior teacher reported that the Deputy School head, who was delegated the 'Chief of Assessment' at school level, once asked: *"what marks were needed by Officers?"* referring to Officers from Examination Body who had come on their supposedly regular spot checks on performance marks. Such a question revealed the extent of lack of awareness of school management on one of their fundamental roles.

The Examination Body should not be spared the blame for lack of monitoring the production of performance assessment marks, resulting in the moderated school-based marks being unauthentic, rendering the outcomes neither valid nor reliable. Monitoring and supervision of the performance assessment should be a system approach involving every process in the system to assure quality. Concentrating on system processes facilitates detection of early process variations resulting in effecting corrective actions timely (Doty, 1996).

Neither school management nor senior teachers who visited teachers during the conduct of assessment to get first hand information on the problems encountered or offered advice and assistance. Lack of supervision internally could be blamed on deficient functional internal policy on submitting performance assessment marks, and the absence of the overarching external performance assessment policy. One senior teacher noted: *“Teachers hold onto marks till a month before moderation. I never ask for marks on regular basis”*. While another senior teacher said: *“I told them to submit marks to me and keep a copy for themselves. But teachers sometimes don’t submit”*.

Analysis of the interviews concurred with the outcomes of quantitative analysis. Raw data was coded and organised into conceptual categories and creating themes. Coding as an integral part of data analysis was guided by research questions. Mechanical reduction of data through open coding yielded initial eleven themes, which came from literature, terms used by respondents, and new thought stimulated by immersion in the data. These were:

(i) Attitudes towards the subject; (ii) Product assessment; (iii) Lack of collaboration; (iv) Insufficient supervision; (v) Lack of standardization; (vi) Inadequate assessments; (vii) Insufficient Training: teachers trained only to teach; (viii) Workload; (ix) Provision of resources; (x) Assessment criteria; (xi) Motivation:

Further analysis (axial coding) through making connections among themes and elaborating the concepts themes represented resulted in collapsing some of the themes. The resulting themes were (i) Product assessment which incorporated inadequate assessment; (ii) Attitude towards the subject, which incorporated contempt towards the subject and motivation; (iii) Monitoring and supervision, (iv) lack of standardisation which incorporated assessment criteria, (v) Resources which incorporated workload. A new theme that emerged through interaction with the data was (vi) plagiarism. Students were said to copy previous students’ projects or giving their relatives to do the projects for them. Some even went to the extent of

buying from the market. After identifying major themes, data and previous codes were scanned to selectively look for cases that illuminate themes (selective coding). These major themes were grounded in the data and formed the major subsequent work of this thesis.

## 5.5 CONCLUSION

The discussion about how assessment is conducted in Botswana senior secondary schools highlighted the extent to which teachers were highly qualified to teach but not to assess. All teachers had at least degree qualification. However, both teachers and school administrators were deficient in skills to conduct assessment in general. Despite that, little in-service training was being conducted to equip teachers with the necessary assessment skills. Induction on performance assessment needs not be confined to school personnel, but to the entire public and students, as important stakeholders in the education system.

Performance assessment was given secondary treatment to standardised testing. All resources were channelled to standardised testing at the expense of performance assessment. For example, performance assessment requires more time which curriculum developers never take into account when allocating the number of periods per subject. If this was done, it would help Department of Secondary Education to rationally allocate realistic loads. As was the case, teachers seemed under worked simply because the assessment aspect had not been factored into their workloads. The provision of tools and equipments for use during practicals was insufficient.

Thorough scrutiny into the workloads revealed that they were relatively high due to high classes of up to 50 students in a class. The high workloads, coupled with lack of resources (physical and time), compelled teachers' assessment to concentrate more on products and artefacts. However, even product assessment was inappropriately conducted. Assessment was teacher-centred, with little opportunity for students to determine what they need to learn. All decisions about their assessment were taken by their teachers, and they were passive recipients. It is well documented gains of active participation by students in their assessment. It also emerged that teachers never standardised their scoring, and each teacher devised his/her own scoring guide and applied it to his/her class.

Due to lack of training by both senior teachers and school administrators in performance assessment matters, supervision was found to be inadequate. Senior Teachers only inspected records and did not visit teachers on site, whilst school administrators' role was minimal. This was so serious to the extent that when moderators arrived to schools they found that performance assessment marks were still with individual teachers. Teachers, school administrators and students had negative attitudes towards performance assessment, hence its treatment as secondary, a view resulting from the subject being used as a filler subject for those students who could not be accommodated by other subjects. The school administration was found to be trying hard to instil a positive outlook on students, but teachers unanimously agreed that the outcome of school-based performance assessment was neither valid nor reliable, and called for an overhaul of the current system.

## 5.6 IMPLICATIONS FOR DESIGN OF INTERVENTION

As concluded in Section 5.5, assessment was characterised, inter alia, by high workloads, large class sizes, tasks of non-equivalent demands, teacher-centeredness, limited time, inadequacy of tools or equipment, negative attitude, and unavailability of standard criteria. All these hindered effective conduct of performance assessment resulting in assessment merely auditing students' learning rather than evaluating learning to direct improvements. All these factors had a bearing on the development of an intervention to improve the validity and reliability of assessment. The development of performance assessment took the form of tasks. Because of the nature of the content, the skills included were different in their demands, hence teachers would choose skills to assess, based on the context and availability of resources.

Initially, tasks were fully described and the criteria for good performance outlined for easy comprehension, thus increasing the validity and reliability. Then the skills assessed for each task and performances to achieve those skills were outlined. The performance criteria outlined (i) the condition under which the performance was to be conducted, (ii) the behaviour to be exhibited, and (iii) the criterion to be fulfilled. These assessment instruments for assessing the tasks and the task itself constituted the assessment materials. They were developed together with practitioners to increase relevance and adoptability. Marks allocation

was explicitly indicated in the assessment instrument, as well as conditions when certain mark(s) was to be awarded to aid teachers to implement assessment in the same way.

In designing the assessment materials, emphasis was placed on multiple assessment of the student using a variety of methods and contexts. Employing multiple assessments was desirable in that it resulted in improved reliability (Airasisan & Russell, 2008; Johnson et al., 2009), just as multiple test items improved the reliability of standardized tests (Rudner, 1994). Since the reduction of class sizes was the jurisdiction of the Ministry of Education Officials, a comprehensible assessment strategy developed has to cater for large class sizes. To make the work less cumbersome and save time using scarce resources, assessment criteria were summarized into a sheet of paper for use in the field. Each sheet of paper could take up to ten students or more. The teacher could objectively score students on a number of skills and make comments for feedback purposes. The scoring made in the field on the summary marksheets was later transferred onto assessment instruments.

Tasks were developed with a view to maintaining a balance between product assessment and process assessment. Some objectives obviously landed themselves to product assessment while others easily landed themselves to process assessment. Agriculture by nature thrives on cooperation, which among students was an important aspect to be incorporated in the development of the assessment materials. However, assessment was individualised, based on what the student did in the group work rather than assigning a group score. Contrary to the notion held by many that performance tasks were easy, developed tasks were such that they were abstract and thought-provoking, and demanded critical thinking and engagement in meaning-making.

The development of tasks also took into consideration the context under which they will be implemented. Because Agricultural activities are highly dependent on weather, tasks were made flexible to fit under various conditions, and since learning entails trial and error, the assessment instrument catered for reassessment. Those who did not get the activities right the first time were given another opportunity to try again. To ensure that tasks were executed as designed, an administration manual was developed alongside the development of the task to guide teachers throughout the country on interpretation. Teachers were trained on how to interpret and implement the tasks. In addition, resources for every task were explicitly delineated. Schools would be required to acquire those resources before they could be

accredited to offer such a task. In other words, schools would be approved to do a certain task after it had satisfied all the necessary conditions, of which resources provision is one of them.

Assessment materials were developed in such a way that they made provision for students to contribute to decisions made about their assessment. In that way, assessment culture that focuses on encompassing evaluation by students of their own learning progress was inculcated (Gasemann 1993). In assessment culture, there is no secrecy (Njabili, 1997; Wiggins, 1998), as the intention is to improve learning by doing rather than audit learning. Both the tasks and assessment instruments were given to students prior to the commencement of conduct of performance tasks. If performance criteria are not given to students they may perform poorly, because they are not aware of teachers' expectation and the criteria for good performance (Airasian & Russell, 2008).

Students prepared themselves prior to the commencement of the tasks, and scored themselves using the assessment instrument during the conduct of the practicals. Such assessment instilled a sense of responsibility among students and was self-monitoring. To authenticate the assessment, both the assessing teacher and Senior Teacher had to endorse their signatures. Once assessment was completed, it was handed to the chief Invigilator for safe keeping, and kept as an official record.

These implications formed the basis of the design of the intervention discussed in the next chapter.

## CHAPTER SIX

### DESIGN, DEVELOPMENT AND EVALUATION OF THE FIRST AND SECOND PROTOTYPES

#### 6.1 INTRODUCTION

The aim of the study is to understand and explore the characteristics and quality processes needed in the performance assessment of Agriculture Form Four students to ensure valid and reliable examinations in Botswana. Findings from Chapter 5 and guidelines from the literature review (Section 3.3) guided the design and development of the intervention as outlined in Section 6.3 taking into consideration the philosophy of learner-centred approach. The literature revealed how performance assessment should be conducted to improve students' performance (Wiggins, 1998). The review of studies conducted in Africa employing design-based research to develop exemplar curriculum materials emphasised the design of the instructional materials (Mafumiko 2006; Tecele 2006; Tilya, 2003), with little or no emphasis on assessment for summative purposes, with the exception of Januario (2008) and Motswiri (2004).

While Motswiri's and Januario's studies investigated assessment, they concentrated on improvement of formative assessment for learning which did not contribute to certification. This study therefore sought to design an assessment intervention which served the dual purposes of improving learning (formative) and evaluating learning (summative). For such an intervention to be widely adopted, it was designed and developed iteratively in collaboration with practitioners and stakeholders (Dick, Carey & Carey, 2009), and involved the development of both the task and the assessment instrument.

In this chapter, product design specifications are outlined in Section 6.2. The description of the first prototype development is presented in Section 6.3, while Section 6.4 delineates formative evaluation of the first prototype. Section 6.5 presents experts views regarding the first prototype. Conclusion of the first prototype is presented in Section 6.6 while implication for design of the next prototype is outlined in Section 6.7. Based on the experts review, conclusions and implications of the first prototype, the design of the second prototype is described in Section 6.8, while its formative evaluation is presented in Section 6.9. The results of the evaluation of the second prototype are presented in Section 6.10. Sections 6.11



and 6.12 delineate the conclusion and implications for the subsequent design of the prototype respectively.

## 6.2 PRODUCT DESIGN SPECIFICATIONS

The baseline survey established the needs for stakeholders in performance assessment (Rainey, 2005), to be integrated in the design and development of quality assurance processes (Abramowich, 2005). It emerged from the baseline findings that there is a need to develop standardised tasks together with their assessment instruments to be used by all teachers throughout the country. When developing tasks, consideration should be given to the current state of resources provision in schools as well as teachers' training in performance assessment. Teachers' workload was high due to large class sizes hence time was an important factor to be factored in the design of the materials. The tasks to be developed should be student-centred to engaged students in the meaning creation of their learning with the hope to improve their negative attitude towards performance assessment. The design also drew heavily on the literature review regarding performance assessment best practices internationally. The conceptual framework provided the roadmap which outlined guidelines and specifications for design of formative evaluation of the assessment prototypes. The product design specifications are delineated below:

### 1. The tasks should be complex and engaging

The tasks developed should address content of importance and substance, and designed in the form of investigations, portfolios and performances, which involved problem-solving that in turn results in report-writing, (Ariev, 2005; Diez, 2002; Macmillan, 2004; Maxwell, 2004; Rennert- Ryan, 2006), rather than just using traditional paper-and-pencil tests (Macmillan, 2004). Such tasks encourage divergent thinking resulting in multiple correct answers to real-world problems. Complex and demanding tasks allow the fulfilment of the primary purpose of improvement in student learning leading to excellence (Wiggins, 1998). Complex tasks would be developed to last longer, encompassing many domains of cognitive, psychomotor and affective skills (Nitko & Brookhart, 2007), for students to use varied multiple skills (Airasian & Russell, 2008; Gardner, 2006).

## 2. Assessment should be integrated into instruction

The traditional standardised testing in which students answer uniform questions in an artificial environment (Wiggins, 1998) are normally designed to audit learning (McMillan 2000; Shepard, 2000), and as such could not be integrated into instruction. School-based assessment of performance tasks has the potential to improve learning if conducted properly and integrated into instruction, as it reveals students' strengths and weaknesses which served as inputs in designing appropriate remedial actions by the teacher. . Performance assessment would be infused in a normal lesson, allowing for assessment to be done when students are ready (Harlen, 2006) and reassessment whenever they did not do well. To assist in having meaningful participation by students, they would be alerted to the teachers' expectations and given assessment materials in advance to familiarise themselves.

## 3. Assessment should be aimed at both processes and products

Product assessment is a crucial aspect of performance assessment, especially where the procedure has been mastered by students (Gronlund, 2006). However, bias towards assessment of the product has the potential to conceal students' capabilities in other domains, such as manipulation. Dick et al. (2009) posit that to determine if learners have achieved an attitude, they have to do something, namely a psychomotor, intellectual or verbal skill. For those skills of which there is only temporary evidence, it is important that one assesses the processes, as it is believed that the repeated use of these improves the product (McMillan, 2004). To effectively assess thinking processes the students undergo in constructing their responses (Airasian, 2005), performance assessments are designed for use under varying contexts, to present all students with the opportunity to showcase their skills. Sometimes it is difficult to prescribe whether to carry out product or process assessment. In such cases it is left to the teacher to use his/her professional judgement as to when to assess product and when to assess the processes to balance the two (see Gronlund, 2006).

## 4. Assessment should be authentic

Student-centred learning takes place in a context in which real life problems manifest themselves in varied forms, and require pragmatic approaches to their solution. Rather than crafting standardised practical tests to be administered to all students throughout the country (Resnick & Resnick, 1992), a variety of authentic tasks would be developed to be applied in

the prevailing context (Johnson et al., 2009; Nitko & Brookhart, 2007). To judge the degree of authenticity of the tasks, Wiggins' (1998:22-24) six standards would be applied:

*Are the tasks realistic?* Do the tasks replicate the ways in which a person's knowledge and abilities are tested in real world situations?

*Do the tasks require judgement and innovations?* Does the student have to use knowledge and skills wisely and effectively to solve unstructured problems, and does the solution involve more than following a set of routine?

*The tasks should involve the students doing something.* The student has to carry out the exploration and work within the discipline of the subject area, rather than restating what was already known or taught.

*Replicates or simulates the context in which adults were tested in the workplace, in civic life, and in personal life.* Do contexts involve specific situations that had particular constraints, purposes, and audiences?

*The tasks should require the student's ability to efficiently and effectively use a repertoire of knowledge and skills to do complex tasks.* Students are required to integrate all knowledge and skills needed, rather than to demonstrate competence of isolated knowledge and skills.

*The tasks should allow appropriate opportunities to rehearse, practice, consult resources, and get feedback on and refine performances and products.*

##### 5. The tasks should be feasible given the resources available in schools

Tasks would be designed to foster collaboration and cooperation to cater for both inadequate time and resources availability in schools (Subsection 5.3.4), as well as large classes (Subsection 5.2.4). Tasks should not only doable but also developmentally appropriate to develop thinking in a variety of ways (Wiggins, 1998). Some would be designed as simulations to serve as an intermediate step to performances that are complicated, involving a higher degree of realism, requiring expensive equipment, or those that put other people's lives in jeopardy (Popham, 2005). For example, instead of children under the age of 16 applying chemicals to crops, they would use water, as they are not legally allowed to use chemicals. Alternatively, they could collaborate with older students.

## 6. Tasks should be evaluated analytically and holistically

Complex performances require that several learning targets or several parts of the performance be assessed using several scoring rubrics consistently (Johnson, et al., 2009) to eliminate subjective scoring (Arter & McTighe, 2001). Some tasks would be crafted to be assessed holistically while others would be assessed analytically. A holistic task is one which is scored using a scale containing several criteria, yielding a single score that gives an overall impression or rating (McMillan, 2004), while a task requiring analytical scoring is one in which each scoring criterion receives a separate score (Nitko & Brookhart, 2007). Because holistic tasks offer little information that can be used for formative feedback, tasks which are scored analytically are inevitable. Analytic scoring separates the whole into parts but takes longer to create and score (McMillan, 2004; Popham, 2005). Each of the holistic and analytic tools would be applied in different situations. The assessment guidelines would be developed simultaneously with tasks (Johnson et al., 2009).

## 7. The assessment should be continuous and cumulative in nature

Evidence of learning is normally collected over time and in the form of a student portfolio (Maxwell, 2004). Assessment would be continuously conducted and integrated in the learning process (McMillan 2000), using multiple methods and raters (McIntire & Miller, 2007; Thorndike & Thorndike-Christ, 2010), with the opportunity for reassessment to approximate the student's true score (Raffan, 2000).

## 8. Self-Assessment

Performance assessment is an open activity (Njabili, 1987), hence assessment materials should be shared with students in advance, as well as parents and the public at large. Nitko & Brookhart (2007) assert that for the students to have their attitudes adequately assessed they should be provided with information about why they should act in a certain way. This helps in shaping an attitude and thereby increasing the chances that desired behaviour will be demonstrated. Having students to assess themselves (Harlen, 2006) helps them to reflect on their performance by applying established criteria to judge their own work. Their self-assessment also helps teachers in formulating sound corrective actions.

## 9. Assessment should produce a traceable evidence of assessing

Retrievable and traceable records should be kept (Le Grange & Reddy, 1998) for assessment to be used for different purposes, and different offices of the Ministry of Education make use of them. To increase the reliability and validity of records, the performance assessment tasks should be designed to lend themselves to minimal record keeping.

### 6.3 DEVELOPMENT OF THE FIRST PROTOTYPE

The description of the tasks that were developed is presented in this section. Tasks development was based on three content areas of *Preparing a plot and planting*, *Applying fertiliser as basal dressing*, and *Controlling weeds using chemicals*.

#### 6.3.1 Description of tasks

The selection of subject content was based on what schools were offering. At the time of conducting this research, schools were offering *Field Crop production*. Tasks from this content were inevitable for the study to operate in synergy with the school's programme, so as to cause minimal disruptions as per the requirement of the Permission Letter from the Ministry (Appendix 4.12). Naturally, the three tasks varied, with *Applying fertiliser as basal dressing* entailing in essence activities with temporary evidence, dictating that assessment involve mainly observations of processes. Meanwhile, *Preparing a plot for planting* and *Controlling weeds using chemicals* involved both observation and product assessment. As a result, it was critical that the development of the intervention captured both.

**Task 1: *Preparing the plot and planting*:** The task has five skills, some of which could not be repeated once they had been performed (Appendix 6.3). Given the class sizes of 35 students (Section 5.2.4), it is difficult to observe all students in these skills, so those who could not be assessed could be assessed in others of similar demand.

**Task 2: *Applying fertiliser as basal dressing*:** The task has six skills which involve mainly assessment of activities (observation) and record keeping (see Figure 6.3 and Table 6.1). There is little or no product assessment and most of the skills could not be repeated once they had been completed. This creates a challenge for the teacher to assess as many students as possible, as the task could be completed in 120 minutes.

Task 3: *Controlling weeds using chemicals*: The task has six skills which involve assessment of activities (observation), the product, and record keeping (Appendix 6.4). This is a typical example of simulation. It could be carried out at any time and repeated as desired; hence its timeframe was not limited.

### 6.3.2 Skills equating

Because of the large classes, it was not possible to assess all students in each skill. To circumvent the problem of assessing students in skills of different demands, skills were equated as shown in Figure 6.1 (below). Using Task 1 as an example, skills 1 and 3 were of equivalent highest demand. Some students could be assessed in skill 1 while others could be assessed in skill 3. Skill 5 was of average demand. Similarly, skills 2 and 4 were of equivalent lowest demand. Skills equating for Task 2 and 3 are presented in Appendix 6.2. Skills equating is a negotiated subjective task which is made more objective by employing more subject matter specialist to individually judge the content and comparing their outcomes to reconcile the discrepancies. The reconciliation process has no hard rules but is premised on mutual discussion and agreement.

Level of demand	Skill				
	1. Preparing a plot (12)	2. Using tools (7)	3. Planting (9)	4. Returning tools & materials to s/room (4)	5. Recording transactions (10)
1					
2					
3					

Figure 6.1: Skills equating for task 1

### 6.3.3 Task Development

The development of the first version of the prototype was undertaken by the researcher guided by findings in Chapter 5, together with the guidelines discussed in Section 6.2. The two main assessment objectives (Objective 2, Assessing the handling and application of information and problem solving skills; and objective 3, Assessing practical and investigative skills (discussed in Section 2.8) guided decisions related to tasks content and design activities.

The task of designing and developing quality intervention began with scrutinizing terminal objectives of the assessment syllabus (Ministry of Education, 2000b). Terminal objectives described exactly what the students should be able to do in a created learning context, not the real world (Dick et al., 2009). Once terminal objectives were comprehended, subordinate skills were derived, these being building blocks to be mastered by students towards achieving a terminal objective (Dick et al., 2009). Detailed tasks were then developed based on the subordinate skills identified.

The discussion of task development was based on Task 2: *Applying fertiliser as basal dressing*. The task development was divided into two parts, namely: (i) The task, and (ii) Assessment Instrument. The task comprised three aspects: The overall task (Figure 6.2); Pictorial presentation of skills (Figure 6.3); and Skills with their performance criteria (Table 6.1). The assessment instrument also comprised three aspects: Scoring instrument - Checklist and Scale (Tables 6.2 and 6.3); Summary marksheets (Table 6.4); and a detailed description of the assessment criteria (Table 6.5). The objective of conducting the practical was delineated, which was *to apply fertiliser as basal dressing*. The task was then stated generally, as presented in Figure 6.2 (below).

Figure 6.2 (below) states only the task, but does not give methodical details of the steps involved in executing the task. This presents implementation problems and is apt to be interpreted differently, resulting in various schools executing tasks of non-equivalent demands. Detailing subordinate skills pictorially in the form of hierarchical analysis shown in Figure 6.3 was a necessary intermediate step in refining the task.



**Given a plot, use appropriate tools from the storeroom to apply a basal dressing fertiliser to your crops.**

**The task will be complete when you have:**

- i. applied the correct fertiliser;
- ii. applied the right quantity of fertiliser;
- iii. used the right tools to measure the quantity of fertiliser;
- iv. applied/observed safety to self, crops and others during fertiliser application to crops;
- v. returned the tools and other materials to the storeroom; and
- vi. recorded all activities carried out.

**Your performance on each step will be judged using the following general criteria:**

- i. performing each step;
- ii. executing each step using the appropriate tools in the proper manner;
- iii. observing safety to self, crops and others all the time;
- iv. keeping detailed records of the activities carried out.

*Figure 6.2: The overall task showing each step and general criteria*



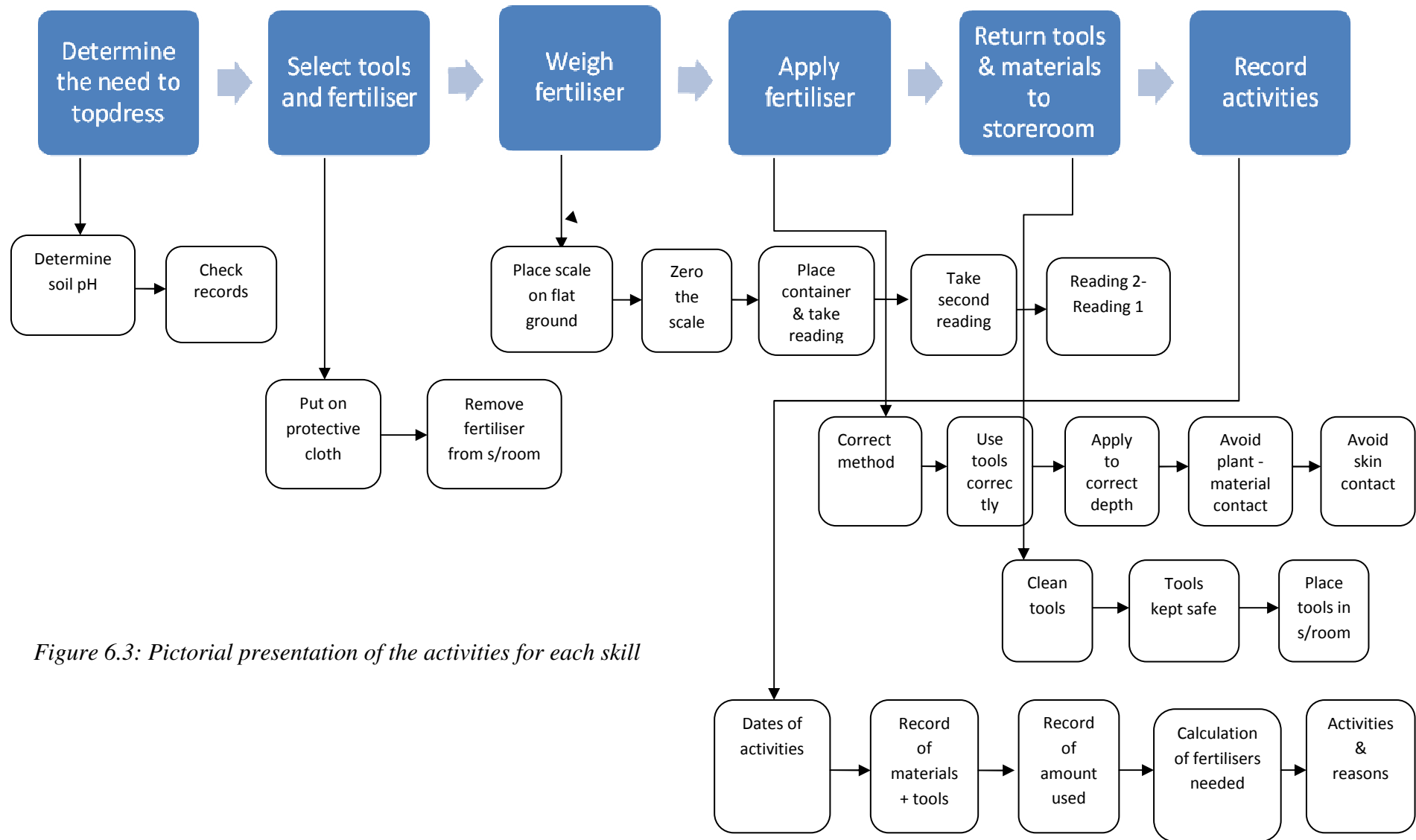


Figure 6.3: Pictorial presentation of the activities for each skill

The subordinate skills and their analysis however were brief for practical purposes. There was a need for increased specificity in the conditions and criteria for performance, as well as prescription of special circumstances (Table 6.1, below), which resulted in performance objectives that guided students as to the precise behaviour expected of them. These included the condition for performance identified using letters CN, the behaviours expected of students identified using letter B, and the performance criteria indicated by letters CR. The condition in this context was the description of the environment, tools and resources that would be available to the learner when performing the skill. The behaviour was the description of the skill that would include actions, content, and concepts, while the criteria were descriptions of acceptable performance of the skill (Dick et al., 2009).

Table 6.1: *Performance skills and matching performance objectives*

Skill	Performance Criteria	Marks
<b>1. Determining the fertiliser requirements (2)</b>	Given plot (CN), determine the need to basal dress (B).	
	<ul style="list-style-type: none"> <li>a. Determine the soil pH,</li> <li>b. Find out what and when fertilisers need to be applied (CR).</li> </ul>	<p>1</p> <p>1</p>

Skill	Performance Criteria	Marks
<b>2. Selecting tools and fertiliser(s) (3)</b>	Given tools and fertilisers stacked in a storeroom (CN), select tools needed and fertiliser for application (B).	
	a. Identify tools and fertilisers needed for application	1
	b. Put on protective clothing,	1
	c. Remove the fertiliser from the storeroom to place of weighing (CR).	1

Skill	Performance Criteria	Marks
<b>3. Weighing the fertiliser (5)</b>	With the fertiliser to be applied and the tools needed ready (CN), weigh the fertiliser (B). <ul style="list-style-type: none"> <li>a. Zero the scale,</li> <li>b. Place an empty container on the scale and take reading 1,</li> <li>c. Place the required amount of fertilisers in the container and take reading 2,</li> <li>d. Subtracting reading 1 from reading 2,</li> <li>e. Work cooperatively with others (CR).</li> </ul>	 1 1 1 1 1

Skill	Performance Criteria	Marks
<b>4. Applying fertiliser (5)</b>	Given the crops growing in a plot (CN), apply the correct amount of fertiliser as basal dressing (B). <ul style="list-style-type: none"> <li>a. Use the correct method of fertiliser application,</li> <li>b. Use tools correctly,</li> <li>c. Apply fertiliser to the correct depth,</li> <li>d. Avoid fertiliser-planting material contact,</li> <li>e. Avoid skin contact (CR).</li> </ul>	 1 1 1 1 1

Skill	Performance Criteria	Marks
<b>5. Returning tools and materials to storeroom (4)</b>	After applying the fertiliser (CN), return tools and materials to the storeroom (B). <ul style="list-style-type: none"> <li>a. Clean all tools,</li> <li>b. Carry tools and materials safely to the storeroom,</li> <li>c. Place tools and materials in the storeroom neatly and in their correct place,</li> <li>d. Work diligently with minimal supervision (CR).</li> </ul>	1 1 1 1

Skill	Performance Criteria	Marks
<b>6. Recording transactions (10)</b>	As you carry out the activities leading to application of fertilisers (CN), record all the transactions carried out and keep a tidy record (B). <ul style="list-style-type: none"> <li>a. Dates of activities,</li> <li>b. Materials and tools used,</li> <li>c. Calculations of amount of fertilisers needed.</li> <li>d. Activities carried out and their reasons (CR).</li> </ul>	1 3 3 3

Numbers in brackets () under the heading Skill are marks for that particular skill

NB: CN = Condition; B = Behaviour; CR = Criteria

Up to this point, the task was considered fully developed. Efforts were now redirected towards developing the accompanying criteria for assessing the task. In developing the most appropriate instrument to evaluate the students learning, a number of factors were taken into account, including (1) the nature and complexity of the elements observed, (2) the time available for: observation, making judgement, and recording judgment, (3) the accuracy or consistency with which the evaluator can make the judgment, and (4) the quality of feedback to be provided to the learners (Airasian & Russell, 2008).

Given conditions prevailing in the assessment arena of Agriculture in Botswana schools (See Section 5.4), one would be obliged to develop an instrument that would be easy to use yet tapping relevant information about the students learning, and resulting in valid and reliable inferences made from assessment scores. A checklist and a rating scale were thus developed (See Tables 6.2 and 6.3). A checklist was used for holistic assessment while a rating scale was used for analytic evaluation of subcomponents of a performance or product (Airasian & Russell, 2008). However, the checklist did not provide enough information for feedback to the students (Dick et al., 2009), and rating scales yielded less reliable scores than checklists (Colton & Covert, 2009). Up to four levels for the rating scales were included for ease of differentiating students' performances. Because of ease of scoring with four levels this helps to improve the reliability of scores.

Table 6.2: *Scoring instrument (Checklist) for the task*

Instructions to the teacher

Score the students using this checklist for skills 1-5 listed in Table 6.1. The scoring is based on ‘yes’ which represents criteria achieved or ‘no’ which represents criteria not achieved. The total mark for each criterion is one. Comment on why the student did not achieve the criteria.

**Student Name:** \_\_\_\_\_ **Score:** \_\_\_\_\_

Skill	Criteria	Scoring Rubric		Comment
		Yes	No	
<b>1. Determining the fertiliser requirements (2)</b>	a. Determine the soil pH			
	b. Find out what and when fertilisers need to be applied			
<b>2. Selecting tools and fertilisers (3)</b>	a. Identifying tools needed for application			
	b. Putting on protective clothing			
	c. Removing the fertiliser from storeroom to place of weighing			
<b>3. Weighing the fertiliser (6)</b>	a. Zeroing the scale			
	b. Placing the container on scale and taking reading 1			
	c. Placing fertilizer on container and take reading 2			



	d. Subtracting reading 1 from reading 2			
	e. Working cooperatively with others			
	a. Using correct method of fertiliser application			
<b>4. Applying Fertiliser (5)</b>	b. Using tools correctly			
	c. Applying fertiliser to the correct depth			
	d. Avoiding fertiliser-planting material contact			
	e. Avoiding skin contact			
	f. Cleaning (if need be) all tools			
<b>5. Returning tools and materials to storeroom (4)</b>	a. Carrying tools and materials safely to the storeroom.			
	b. Placing tools and materials properly in the storeroom.			
	c. Working diligently with minimal supervision			

Teacher's Name \_\_\_\_\_

Teacher's Signature \_\_\_\_\_

Date \_\_\_\_\_

Snr Teacher's Name \_\_\_\_\_

Snr Teacher's Signature \_\_\_\_\_

Date \_\_\_\_\_



Table 6.3: *Scoring instrument (Scale) for the task*

Instructions to the teacher

Score the students using this scale for Skill 6 listed in Table 6.1. The scores range from 0 to 3. Put the student score in the column ‘Mark’.

Comment on why the student did not achieve the criteria.

Student Name: \_\_\_\_\_

Score: \_\_\_\_\_

Skill	Criteria	Scoring Rubric				Mark	Comment
		0 mark	1 mark	2 mark	3 mark		
<b>6. Recording transactions (10)</b>							
	a. Record the date of activities	0-90% recorded	>90% recorded				
	b. Record materials and tools used in each activity	0-50% recorded	50-70% recorded	70-90% recorded	>90% recorded		
	c. Record of calculations of amount of fertilisers needed	All calculations wrong	30-70 % calculations correct	70 -90% calculations correct	>90% correct calculations correct		
	d. Record of activities and reasons	0-50% recorded	50-70% recorded	70-90% recorded	>90% recorded		

*NB: A black box means no mark is allocated*

Teacher's Name \_\_\_\_\_

Teacher's Signature \_\_\_\_\_

Date \_\_\_\_\_

Snr Teacher's Name \_\_\_\_\_

Snr Teacher's Signature \_\_\_\_\_

Date \_\_\_\_\_

However, the above criteria would be very cumbersome for teachers to carry to the field (practical site), because they would have to carry 35 copies (average class size) of each of the two instruments (Tables 6.2 and 6.3 for checklist and rating scale respectively), and frequently paging through to locate the student under observation to award marks. To avoid this problem and simultaneously seeking to conduct quality scoring, a summary marksheet is used instead. Using the summary marksheet, the teacher would need to carry:

- (i) one copy of the class list populated for each student, with letters corresponding to the criteria for each subordinate skill (See Table 6.4).
- (ii) one copy of each of the marking criteria for reference (Tables 6.2 and 6.3).
- (iii) a copy of the detailed description of marking criteria (Table 6.5).

Detailing the marking criteria helps teachers interpret the criteria in a similar way, for example, if one does not have a clear understanding of what is meant by *Use correct method of applying fertiliser (4(a))*, one can find detailed examples of methods of application of fertiliser in the detailed description (Table 6.5).

Table 6.4: *Example of summary marksheet for use by teachers*

**Instructions to the teacher**

Use this marksheet in the field to assess the students. This should be used in conjunction with Table 6.5 which gives a detailed description of each criterion. The letters (e.g. a, b, etc) represent criteria corresponding to each skill as detailed in Table 6.4. Circle the letter when the student meets the criterion corresponding to that criterion. Table 6.6 shows how this marksheet should be completed. **Total marks = 29**

Student Name	Skill						Marks	Comment
	1. Determining the fertiliser requirements (2)	2. Selecting tools and fertiliser (3)	3. Weighing the fertiliser (5)	4. Applying the fertiliser (5)	5. Returning tools & materials to s/room (4)	6. Recording transactions (10)		
Lorato Menwana	a b	a b c	a b c d e	a b c d e	a b c d	a 0 1            b 0 1 2 3 c 0 1 2 3      d 0 1 2 3		
Lingani Mesho	a b	a b c	a b c d e	a b c d e	a b c d	a 0 1            b 0 1 2 3 c 0 1 2 3      d 0 1 2 3		
Tate Mgadla	a b	a b c	a b c d e	a b c d e	a b c d	a 0 1            b 0 1 2 3 c 0 1 2 3      d 0 1 2 3		
Ernest Forbes	a b	a b c	a b c d e	a b c d e	a b c d	a 0 1            b 0 1 2 3 c 0 1 2 3      d 0 1 2 3		
Nonny Meshack	a b	a b c	a b c d e	a b c d e	a b c d	a 0 1            b 0 1 2 3 c 0 1 2 3      d 0 1 2 3		
Bonang Ketshabile	a b	a b c	a b c d e	a b c d e	a b c d	a 0 1            b 0 1 2 3 c 0 1 2 3      d 0 1 2 3		

Teacher's Name \_\_\_\_\_ Teacher's Signature \_\_\_\_\_ Date \_\_\_\_\_

Snr Teacher's Name \_\_\_\_\_ Snr Teacher's Signature \_\_\_\_\_ Date \_\_\_\_\_

Table 6.5: Detailed description of marking criteria for use during field evaluation

Skill	Criteria	Mark
<b>1. Determining the fertiliser requirements (2)</b>	<ul style="list-style-type: none"> <li>Determine soil pH: different crops thrive best in different soil pH. Find out the soil requirements for the crop you are growing, particularly its pH. The teacher should not tell students the type of soil and pH needed. Students should take a leading role in their learning. This criterion is marked from the students' records.</li> </ul>	1
	<ul style="list-style-type: none"> <li>Find out what and when fertilisers need to be applied: The student should check from literature which fertilisers and how much is used as basal dressing for the variety of the crop planted. Note that fertiliser needs for the same variety of crop may differ from one region to the other depending on influence of the climate. This criterion is marked from students' records.</li> </ul>	1
<b>2. Selecting tools and fertilisers (3)</b>	<p>a. Identifying tools and fertiliser needed for application: tools that would be needed are such as scale, hand trowel, while fertilisers are such as N:P:K 2:3:2 (34), superphosphate, kraal manure, potassium sulphate, wood ash (K), basic slag (P), and so on. There could be as many or as few as possible, depending on your location. Prescription of how many tools and/or materials that warrant a point is also dependent on your situation. It is left to the professional judgement of the teacher to determine how many marks the student deserves for the tools and materials enumerated.</p>	1
	<p>b. Putting on protective clothing: Schools should make concerted effort to acquire protective clothing for students and teachers, such as overall, boots, mask, goggles, and respirator. It is a legal requirement that chemicals should be applied putting on protective clothing. This is scored on all or nothing basis.</p>	1
	<p>c. Removing the fertiliser from storeroom to place of weighing: this is a collaborative activity. The teacher should look at what students do when loading and off-loading; whether they are working together as a team. Team building is an important aspect for productivity. Aspects of attitudes are also encompassed.</p>	1
<b>3. Weighing the fertiliser (5)</b>  Group work is expected	<p>a. Zeroing the scale: whether digital or .analogue scales should be zeroed to get precise reading. The scale should be at 0.00 before putting anything on it. Putting the scale on a flat area will facilitate achievement of such.</p>	1
	<p>b. Placing an empty container on the scale and taking its reading 1: (see worked example)</p>	1
	<p>c. Placing the desired amount of fertiliser in the container and taking reading 2 (see worked example)</p>	1
	<p>d. Subtracting reading 1 from reading 2. (see worked example in teacher's guide)</p>	1
	<p>e. Working cooperatively with others: cooperation is highly encouraged in Agriculture. Almost all agricultural activities require group work. It should be inculcated into students that helping one another is an important attribute for success. Competition should be discouraged by all means.</p>	1
<b>4. Applying Fertiliser (5)</b>	<p>a. Using correct method of applying fertiliser: methods such as broadcasting, drill, banding, foliage, etc. Students should justify the choice of their method.</p>	1
	<p>b. Using tools correctly: tools last longer if used for the right purpose. Students should use tools for their rightful purposes. However, this does not relegate improvisation whenever necessary.</p>	1
	<p>c. Applying fertiliser to the correct depth: Fertilisers should be applied at the correct depth to be used by plants. If applied too deep, it will leach resulting in stunted growth by plants. If shallowly applied, it will volatilise and escape into the air.</p>	1

Skill	Criteria	Mark
	d. Avoiding fertiliser-planting contact: Chemicals and fertilisers burn crops. The fertiliser should be placed deeper than the planting material or away from the row of planting material.	1
	e. Avoiding skin contact: fertilisers are chemicals and have residual effect, as such, contact with the skin should be avoided.	1
<b>5. Returning tools and materials to storeroom (4)</b>	a. Clean all tools: Tools should be cleaned before they are taken to the storeroom. Cleaning does not imply only using water. Greasing, polishing or removing soil may constitute cleaning. Cleaning is any action that prevents tools/equipments from rusting.	1
	b. Carrying tools and materials safely to the storeroom: tools should be carried safely to the storeroom to prevent injuries. The proper way to carry tools is pointing the sharp end downwards. Tools left lying around are exposed to harsh weather conditions and are likely to wear quickly.	1
	c. Placing tools and materials correctly in the storeroom: tools should be placed on tool racks. It is advisable that such be made in case they are not there, otherwise it would be difficult to assess this skill.	1
	d. Working diligently with minimal supervision: Students should not work only when the teacher is around. They should take a leading role in their learning. Once students have been cultured into doing right things all the time, they work on their own with minimal supervision. Tools won't be left lying around, they will always submit their record books at the right time for scoring, and willing to share tools, and so on.	1
<b>6. Recording transactions (10)</b>	a. Recording the date of activities: This is marked out of 1 because it does not require any skill to do yet it is important.	1
	b. Recording materials and tools used in each activity: The materials and tools used in each activity will vary in terms of number, type and extent of usage from school to school, activity to activity, and student to student. The professional judgement of the teacher is called for in this particular case.	3
	c. Recording calculations of amount of fertilisers needed: Before going to weigh the fertiliser to apply to crops, the student has to calculate how much fertiliser is needed. All the calculations have to be shown for the student to score maximum points.	3
	d. Record activities and reasons for carrying out the activities: This will depend on individual schools, classes, and students. It is difficult to state categorically the number of reasons that warrant 1 or 2 marks. The teacher is better placed to know. Professional judgement should again be exercised.	3
<b>TOTAL</b>		<b>29</b>

*NB: The total marks are indicated beside each skill in brackets ( ) in the first column and marks for each criterion in the last column.*

Given the summary marksheet presented in Table 6.4 and the detailed description of marking criteria presented Table 6.5, the teacher can assess a number of students with ease. Let us consider how the summary marksheet is used. Taking Lorato Menwana as an example, under the activity *Determining the fertiliser requirements*, there are two criteria labelled *a* and *b*. From Tables 6.1 and 6.2, it was evident that:

criteria *a* is:                   determining soil pH

and criteria *b* is:            finding out what and when fertilisers need to be applied.

The teacher circles *a* and *b* if the student successfully achieves these criteria. Table 6.6 (below) shows how the scoring is done. Other skills are scored the same. The total mark for *Lorato Menwana* was obtained by aggregating the number of letters and/or numbers circled, totalling 22 out of 29, and captured under the column ‘total’. Likewise, the mark for *Lingani Mesho* was computed in a similar way (23).

This kind of assessment infused flexibility and the teacher could assess different students at the same time on the same activity or skill, or assess different activities. Practically, the teacher could target a few students per lesson, perhaps 10, who could be quickly and accurately assessed. The rest of the class can be assessed on other skills of similar demands identified during skills equating. Any criteria not achieved by the student can be reassessed another time or during the conducting of a similar task. After assessing, the teacher transferred the marks from the summary marksheet to the scoring instruments shown in Tables 6.2 and 6.3 (above) for purposes of traceable and retrievable records for accountability. Supervisors and inspectors can peruse these records and use them for reconciliation whenever there is a dispute.

The development of the other two tasks followed a similar procedure, and is presented as appendices 6.3 and 6.4.

Table 6.6: Sample of completed summary marksheet

**Total: 29 marks**

Student Name	Subordinate skill						Total	Comments
	1. Determining the fertiliser requirements (2)	2. Selecting tools and fertiliser (3)	3. Weighing the fertiliser (5)	4. Applying fertiliser (5)	5. Returning tools & materials to s/room (4)	6. Recording transactions (10)		
Lorato Menwana	(a) (b)	(a) b (c)	(a) (b) d (c) (e)	(a) (b) d (c) (e)	(a) (b) c d	a 0 (1) b 0 1 2 (3) c 0 1 (2) 3 d 0 (1) 2 3	22	
Ernest Forbes	a b	a b c	a b c d e	a b c d e	a b c	a 0 1      b 0 1 2 3 c 0 1 2 3    d 0 1 2 3		
Lingani Mesho	a      (b)	(a) (b) (c)	(a) (b) (c) d (e)	(a) (b) d (e)	(a) (b) (c)	a 0 (1) b 0 1 2 (3) c 0 1 (2) 3 d 0 (1) 2 3	23	
Nonny Meshack	a b	a b c	a b c d e	a b c d e	a b c	a 0 1      b 0 1 2 3 c 0 1 2 3    d 0 1 2 3		

Teacher's Name \_\_\_\_\_

Teacher's Signature \_\_\_\_\_

Date \_\_\_\_\_

Snr Teacher's Name \_\_\_\_\_

Snr Teacher's Signature \_\_\_\_\_

Date \_\_\_\_\_

## 6.4 FORMATIVE EVALUATION OF THE FIRST PROTOTYPE BY EXPERT GROUP

The preliminary validity of the prototype was to be ascertained through evaluation and feedback could be incorporated into the redesign of the second prototype.

### **6.4.1 Research Design**

The evaluation of the first prototype of the standard task and assessment materials was carried out by experts who were given the three tasks to review against the criteria. The intention of evaluation at this stage of development was to maximise the content validity (De Villis, 2003) as well as the consistency between tasks and assessment criteria (Plomp & Nieveen, 2007). Feedback from evaluation was incorporated into the development process to improve the effectiveness of the intervention (McDavid & Hawthorn, 2006). The review was guided by the evaluation question:

*Is there consistency between the tasks and assessment criteria?*

### **6.4.2 Participants**

Formative evaluation of the first prototype involved two groups of evaluators, the first of which comprised three Assessment Officers, two Agriculture Education Officers from CD&E and DSE, and a lecturer from the College of Agriculture offering Measurement courses to student-teachers. Officers also had wide experience of the school system as they had been teachers before. The second group of participants comprised five teachers from two senior secondary schools who had vast experience in teaching Agriculture. Demographic information relating to the experts is given in Appendix 6.5.

### **6.4.3 Data collection strategies**

Expert Evaluators completed 4-point Likert scales ranging from 4 representing strong endorsement, to 1 representing weak endorsement, as shown in Appendix 6.6. The first scale sought to find out the sufficiency of the format and the clarity of the language. The next two scales were aimed at determining the adequacy of the task and adequacy of assessment criteria. The scale sought to find out the clarity of instructions and the instrument also provided experts with the opportunity to express their views by answering open-ended questions. Quantitative data was analysed descriptively through determining the reliability of



the instrument. Data was analysed qualitatively using themes and thick descriptions quoted verbatim.

## 6.5 EXPERTS' VIEWS AND EXPERIENCES WITH THE FIRST PROTOTYPE

Experts were requested to evaluate the *Task and Assessment Instrument* for (i) sufficiency of format, (ii) clarity of language, (iii) adequacy, and (iv) clarity of instructions, on scales ranging from 1 to 4. Negatively worded items were reversed during analysis so that the high numbers denoted high endorsement. Table 6.7 presents the number of items and reliability coefficients for each scale. Detailed table is presented in Appendix 6.7.

Table 6.7: *Reliability coefficients for scales of the tasks and assessment instruments*

Scale name	No of items	Reliability coefficient of task		
		1	2	3
Sufficiency of format of task	3	.82	.64	.91
Sufficiency of format assessment instrument	5	.89	.94	.90
Clarity of language for task	3	.93	.82	.92
Clarity of language for assessment instrument	5	.92	.96	.88
Adequacy of task	10	.83*	.95	.87
Adequacy of Assessment Instrument	7	.77	.87	.90
Clarity of instructions	4	.90	.88	.87

\* when 1 item is removed

### Task 1: Preparing a plot and planting

The reliabilities of the sub-scales for task 1 were found to be generally high as shown in Table 6.7. The Generally, there was high degree of agreement among experts concerning the sufficiency of *format of the* task and assessment instrument, adequacy of the task and assessment instrument for Form Four level, and clarity of the language used. However, the

task was found to be physically inaccessible and lacking activities targeting students' of different abilities:

*Assessment task did not cater for those who are disabled or visually impaired even though they are mixed in the classes with normal students. Their abilities are not assessed because most of the work is done for them by other students or assistant teachers.*

Experts also suggested separating this task into two tasks since *Plot preparation* and *Planting* were adequate enough to stand on their own. Suggestions were made to improve clarity of the instructions. Experts' views are summarised in Table 6.8.

Table 6.8: *Summary of experts views on task 1*

<b>critterion</b>	<b>Experts views</b>
<b>Overall quality of task</b>	<p>The task is of good quality and achievable. The assessment instrument is objective and allows the one assessing to spread out the marks across a wide spectrum of performance by the learners. It can be useful for narrowing the gap in mark allocation by different examiners in different places since the instruments clearly elaborate on what must be done in order to award each mark.</p> <p>Land preparation and planting skills be separated. Marking criteria should be differentiated to show that the tasks cater for learners of different abilities.</p>
<b>Content of task</b>	<p>Covers much of the content of the syllabus</p> <p>There is no description of how content of the task relates to the curriculum objectives.</p> <p>In some areas demanding tasks given low marks and level of difficulty should correspond with mark allocated.</p>
<b>Format of task</b>	<p>Well formatted and structured to measure the indented learning outcome. The structure is technically sound and of technical quality.</p> <p>The structure should be such that it shows the relationship between the task and assessment. The task is not differentiated to accommodate the different ability levels. Needs improvement, it is too congested and shorten sentences</p>

---

<b>Language of task</b>	Simple to be understood by learners at the proposed level. Language is ok but instructions not quite clear. Guidance not quite clear
-------------------------	--

---

### Task 2: Applying fertiliser as basal dressing

All the sub-scales were found to be internally consistent with generally high reliabilities as shown in Table 6.7 (above). Reliability for *Adequacy of assessment Instrument* was .87 after one item was removed. Experts unanimously agreed that the standard task and assessment materials were well constructed to elicit the desired outcomes.

Experts agreed that the task and assessment instrument were sufficiently formatted and the language clear enough to be understood by the intended users: “*The task is well framed and phrased for one to easily follow and understand and I will suggest you retain them in this format*”. Likewise, experts agreed that the task and the assessment instrument were adequate to be used for Form Four level to measure students’ capabilities more efficiently than before, with the exception of not being physically accessible to all students. Experts felt that the task was not providing opportunities for all students to interact and cooperate within a group. . The only concern raised was that it seemed inadequate in assessing students’ cognitive skills. Table 6.8 summarises experts’ views for task 2.

Table 6.9: *Summary of experts views on task 2*

<b>critterion</b>	<b>Experts views</b>
<b>Overall quality of task</b>	<p>Clearly, well defined, explicit, good, elaborated and clearly outlines the actual activities that occur during fertiliser application for easy understanding by both teachers and students. Comprehensive and user-friendly to execute.</p> <p>I haven't been able to come up with a document/planned assessment instrument to use, now light has been shed.... just wish the instrument could be used on a small sample of students to see if it is user friendly.</p> <p>Instructions to the tasks not clear particularly for the pictorial presentation.</p>
<b>Content of task</b>	<p>Good, self-explanatory, well understood, fig 6.4 skill 6 criteria (b) and (c) will require more details, covers all content areas for the level of intended users.</p> <p>Content does not show any differentiation among the students' of different abilities. Activities need to be reduced.</p>
<b>Format of task</b>	<p>Well structured, detailed, clear and easy, too congested &amp; a lot of wording.</p> <p>The structure does not present itself to allow interaction among learners.</p>
<b>Language of task</b>	<p>Comprehensive, and clear but shorten writing, easy to understand and suitable to both students and teachers.</p>

### Task 3: Controlling weeds using chemicals

The reliabilities of the sub-scales for task 3 were found to be high as presented in Table 6.7 (above). Experts expressed satisfaction with the format, language and adequacy of task 3. However, suggestions were made to improve the structure so that it would differentiate between students of different abilities, as well on its feasibility given the nature of resources involved in this particular task. *The task on weed control using chemicals is well described but I am in doubt of its feasibility in a school set-up particularly taking into account the residual effect of herbicides in a garden with multiple users.* Table 6.10 summarises experts' views on task 3.

Table 6.10: *Summary of experts views on task 3*

<b>critterion</b>	<b>Experts views</b>
<b>Overall quality of task</b>	<p>Hope this exercise will be given the seriousness it deserves because at the moment or for a very long time agric practicals haven't been well assessed. Opinions of different bosses have been used as assessment tools. Well described, but its feasibility in a school is doubtful taking into account the residual effect of chemicals.</p> <p>Task composition is mainly suitable to science based students with better knowledge of mathematics.</p> <p>Need to have a set standard of tasks for each syllabus topic and common for all schools</p>
<b>Content of task</b>	<p>Covers content of syllabus and relevant. Covers all activities in the task.</p> <p>If there are no weeds, shouldn't the learner choose an appropriate method s/he would like to control the weeds?</p>
<b>Format of task</b>	<p>Well structured and easy to follow. Measure what is intended to measure.</p> <p>I have learnt the detailed step by step approach for assessing tasks and the benefit of using a well prepared scoring instrument to assess and grade students. Improve the format</p>
<b>Language of task</b>	<p>Clear and simple</p>

## 6.6 CONCLUSION

There was high endorsement on the need to develop the assessment intervention to improve the assessment of practicals and consequently enhance its contribution for certification. Expert welcomed the development of the standard task and assessment materials as the solution to some of the problems that had bedevilled practical assessment for a long time.

Almost all experts expressed lack of understanding on the use of the marksheets, calling for a thorough training of teachers before they could be given the assessment instruments to use. The tasks were found to lack i) physical accessibility to all students; ii) demonstration by students of understanding in a variety of ways; iii) activities for groups of differing abilities;

iv) opportunities for all students to interact and cooperate within a group set up, and (v) assessment of affective skills. Suggestions were made on how to include these.

Other suggestions made to be factored in the design of the second prototype included improving clarity of instructions, and inclusion of date of task execution as authentication, as well as teacher's and senior teacher's signatures as a quality assurance step, justification for high marks for record keeping as they lacked face validity.

## 6.7 IMPLICATIONS FOR FURTHER DEVELOPMENT

Findings of the experts' evaluations resulted in tasks being revised to make them open and more interactive, both between students and between students and teachers. Interaction helps students to learn from multiple sources in different ways, as well as use information to evaluate themselves and others.

Since experts indicated that tasks did not cater for special needs students, expertise was sought from the Botswana Examinations Council's Special Education Officers on how to craft tasks and assessment instruments so as to cater for special needs students. Tasks were refined to cater for different abilities and differentials in intellectual development. In addition, tasks were made to cover in-depth challenging content of knowledge and skills (Lane & Stone, 2006) to provide students with the opportunity to use multiple skills and abilities (Diez, 2002; Rennert-Ariev, 2005; Ryan, 2006).

Instructions were revised so as to be easily understood by all users, a crucial move since assessment materials would be given to students prior to teacher assessment for familiarisation and self-evaluation. A number of affective objectives would be introduced as they are an integral part of the curriculum, and are no longer considered part of the 'hidden curriculum' as was the case during the 1980s (Jarolimek, 1981). However, they would be infused in skills assessment (Gronlund, 2003) for ease of assessment.

## 6.8 DESIGN OF THE SECOND PROTOTYPE - PILOT

The intention of design and evaluation at this stage was to explore the validity and practicality of the standard task and assessment materials in the context of Botswana Agriculture practical assessment with Form Four students. That is, to find out if teachers were able to use the standard task and assessment materials with their students as intended by the designer. The design and review were guided by the evaluation question:

*What is the practicality of the intervention that aims at supporting performance assessment in agriculture?*

The review of the first prototype was the first cycle of formative evaluation and highlighted a number of issues to be included in the design of the second prototype, as indicated below:

1. Incorporation of collaborative activities: initially, collaborative activities were implicitly infused in the task. These had to be made explicit to guide teachers and students precisely to the behaviour expected to improve the validity and reliability of the assessment.
2. Clarity of instructions: in a number of cases, instructions were improved for ease of understanding. An *introduction* and *Directions to teachers* were included to guide teachers and students to the interpretation of the intervention. The *Detailed description of the marking criteria* section of the standard task and assessment materials was elaborated to clarify how each objective could be achieved and scored. The *Implementation plan* was also introduced, detailing how each skill should be implemented.
3. Inclusion of tasks catering for different abilities: the design of the prototype included a number of critical thinking, abstract, problem-solving, and reasoning skills to cater for students with different abilities. These enabled teachers to differentiate between students who needed help and those who needed more challenging work.
4. Inclusion of affective skills: a number of affective skills in the development of the second prototype were included, for instance: (a) Working diligently with minimal

supervision; (b) Working cooperatively with others; and (c) Observing safety to self and others.

5. Marks allocation: experts had raised questions on the allocation of marks, for example why more marks were allocated to record keeping. Efforts were made to explain in the *Detailed description of the marking criteria* why this was the case.
6. Accommodation for special needs students: tasks were made flexible so that they could be easily modified for special needs students.
7. Resources needed: for ease of preparing for the lesson, materials, tools and other resources needed for each task were outlined. This would enable schools to acquire them in advance for effective implementation of tasks.
8. Induction of teachers: well-thought out notes were prepared for use during workshops with teachers on how to implement the intervention. This was particularly important to facilitate standard implementation of the intervention throughout the country.

## 6.9 FORMATIVE EVALUATION OF THE SECOND PROTOTYPE

This section discusses the research design employed in evaluating the intervention, the participants involved, and data collection strategies.

### **6.9.1 Research design**

To ascertain consistency or logical design of the standard tasks and assessment materials, three teachers and their students from one school were involved in piloting the standardised materials. Purposive sampling technique was employed for both the teachers and students, having agreed to voluntarily participate through a signed consent form. Each teacher piloted one of the three tasks and assessment instruments.



### 6.9.2 Participants

Three teachers and their Form Four students from one Government-Aided school participated in the study. There were four teachers in total, three of whom volunteered to participate. Demographic information about the participants is presented in Table 6.11 (below). All teachers were male, with adequate experience of teaching and held Senior Teacher's Grade II position. They possessed at least a bachelor's degree qualification, and their class sizes were large, ranging from 42-45.

Table 6.11: *Demographic information of participants*

Variable	Teacher 2	Teacher 1	Teacher 3
Age	38	34	40
Sex	Male	Male	Male
Academic qualification	Degree	Degree	Degree
Professional qualification	BSc. Agric Ed	BSc. Agric Ed	BSc.Agric Ed
No of year teaching	15	7	14
Post	Ag. ST Grade I	ST Grade II	ST Grade II
Class size	45	45	42
Sampled students	10	5	5

A total of 20 Form Four students participated in the study, ranging in age from 16-18 years. As indicated above, Form Four students were used because the final students (Form Five) had already completed all their practicals. During the time of conducting the study, the Form Fives were preparing a report for the final project (see Sections 1.3 and 2.8) to be ready for scoring and moderation in September. Form Fours were considered to have enough experience in conducting practicals since they were did them at junior level where Agriculture was offered as a core subject (See Section 2.5.3).

### 6.9.3 Data collection strategies

#### (i) Procedure

Teachers were taken through the standard task and assessment materials step-by-step prior to implementation, during the workshop conducted in the afternoon when they had no lessons. However, they had difficulty in conceptualising the implementation strategy, particularly the

*Summary marksheet*, so further efforts were made to explain, reinforced by hands-on facilitation for better conceptualisation. Teachers were subsequently requested to explain to their students how the materials were used. Teacher 1 selected task 3: *Controlling weeds using chemicals*, teacher 2 selected task 1: *Preparing a plot and planting*, while teacher 3 chose task 2: *Applying fertilisers as basal dressing*, as shown in Table 6.12.

Table 6.12: *The tasks selected by teachers*

<b>Teacher</b>	<b>Task</b>	<b>Task Name</b>
1	3	Controlling weeds using chemicals
2	1	Preparing a plot and planting
3	2	Applying fertilisers as top dressing

Materials were delivered to the teachers four days before the planned day of implementation for distribution to students before the start of the lesson. The conduct of the practicals was made during the last week of the term after end-of-term examinations. Though teachers were busy marking examinations, they managed to accommodate the piloting because they understood its likely implications for improving students' learning. During the day of piloting I went early to the school to assist with logistical issues to ensure proper implementation. All the three practical lessons were conducted the same day during normal teaching time.

The researcher accompanied teachers to the garden, where the practicals took place and the teachers introduced me to the students so that I did not appear as a stranger. They gave a brief explanation of the objective of the practical before the students could work on their own. Teacher 1 selected ten students, while teacher 2 and 3 selected five students each for assessment. Task 1 was relatively easy to assess, hence more students were selected. Teachers interacted with the students as they were assessing their processes. Meanwhile, the researcher completed an observation schedule, and at the end of the lesson, teachers completed a self-administered questionnaire to reflect on how they perceived the practicality of the standard task and assessment materials. The researcher requested six students in all from the three classes to be interviewed in the afternoon as a focus group. After the students'

interviews, one teacher was interviewed and the other two were interviewed the following day. All the interviews were audio-taped.

*(ii) Data collecting instruments*

Teacher evaluation questionnaire and interview

At the end of the lesson, each of the three teachers completed a questionnaire which sought their views on the implementation of the standard task and assessment materials, with both closed-ended and open-ended questions. The former targeted teachers' views on instructional behaviour, knowledge of assessment, standardising assessment and class management, while the latter sought the views of teachers on quality, content, format and language used on the standard task and assessment materials.

A structured interview was administered at the end of the lesson, the aim of which was to capture respondents' views about the impression of the intervention in their own words. Issues discussed ranged from the usefulness of the standard task and assessment materials, its feasibility, things they did not like, and how the tasks could be improved. The teacher evaluation questionnaire and interview schedule are presented as appendices 6.8 and 6.9 respectively.

Lesson observation

All lessons were observed and the researcher posed as a silent observer. An observation schedule was used to collect data of the activities of the lesson. The observation focussed on instructional behaviour of teachers, knowledge of assessment by the teacher, and resources availability. The instrument was developed in the form of a rubric to fully describe the activities of the lesson. The *Instructional behaviour* scale had five-level descriptors which holistically described the behaviour of the teacher. The teacher's knowledge and resources availability were evaluated through four-point and five-point analytic scales respectively. The instrument had provision for field notes to capture what transpired during the lesson. The observation schedule is presented as appendix 4.5.

Student interview

A focus group of six students was interviewed at the end of the practical lesson, using a semi-structured interview schedule, with probing and follow-up questions to get insight into

students' views about the standard tasks and assessment materials they had been implementing. The interview lasted for about one hour and transactions were audio-taped for later transcription. The interview schedule is presented as appendix 4.6. Students completed a self administered questionnaire comprising a Likert scale with 11 items and five open-ended questions. The questionnaire is presented as appendix 4.8.

## 6.10 RESULTS OF THE EVALUATION OF THE SECOND PROTOTYPE

Results presented are based on lesson observation, standardising marking, students' understanding of assessment practices, completion of the assessment instrument, and record keeping. Lesson observation is divided into instructional behaviour, knowledge of assessment and resources availability.

### *6.10.1 Lesson Observation*

Lesson observations were conducted with the view to understand what the teacher was doing during performance assessment. Lesson observations constituted instructional approach adopted by the teacher, teachers' knowledge of assessment, and resources availability.

#### Instructional approach

Table 6.13 shows the extent to which teachers' conducted activities to facilitate effective learning. Generally, teachers' instructional practice in assessment was average to above average indicating that their instructional approaches were student-centred. Figure 6.4 (below) shows the frequencies of instructional approaches diagrammatically.

At the beginning of the lesson, students had been given copies of the standard task and assessment materials just before the lesson. Teachers' introduction of the lesson's objective was not clear to the students as they did not understand what to be achieved at the end of the lesson. Teachers asked a number of questions to gauge the students' knowledge on the topic, but that was not related to everyday life. Only teacher T<sub>3</sub> made some effort.. Teachers interacted with the students as they were working and their assessment was individualised. Teachers generally agreed that the intervention was very useful for instructional effectiveness.



*Figure 6.4: The occurrence of the extent of instructional behaviour*

Table 6.13: *The extent of conducting activities by different teachers*

Activity	To little or no	To some extent	To a moderate extent	To a great extent	To a very great extent
1. Stating the performance assessment objective before the start of the practicals			T <sub>3</sub> T <sub>1</sub>	T <sub>2</sub>	
2. Asking questions to gauge the students level of knowledge of the activity		T <sub>2</sub> T <sub>3</sub>		T <sub>1</sub>	
3. Linking the relevance of the practical to everyday lives			T <sub>2</sub> T <sub>1</sub> T <sub>3</sub>		
4. Clarifying what resources are to be used and how they are to be used			T <sub>1</sub> T <sub>3</sub>	T <sub>2</sub>	
5. Clarifying what will be assessed and how			T <sub>1</sub> T <sub>3</sub>	T <sub>2</sub>	
6. Stressing observation of safety			T <sub>2</sub> T <sub>1</sub> T <sub>3</sub>		
7. Spelling out observable aspects of performance that should be judged			T <sub>2</sub>	T <sub>1</sub> T <sub>3</sub>	
8. Distributing the task to the class before the start of the practicals			T <sub>3</sub>		T <sub>2</sub> T <sub>1</sub>
9. Managing time well			T <sub>2</sub> T <sub>3</sub>	T <sub>1</sub>	
10. Organising materials for the practical			T <sub>3</sub>	T <sub>2</sub> T <sub>1</sub>	
11. Stating the students' role				T <sub>2</sub> T <sub>1</sub> T <sub>3</sub>	
12. Stating the students' task				T <sub>2</sub> T <sub>3</sub> T <sub>1</sub>	
13. Emphasising reasoning as opposed to rote learning	T <sub>2</sub> T <sub>3</sub>	T <sub>1</sub>			
14. Providing an appropriate setting to elicit and judge the performance or product				T <sub>3</sub>	T <sub>2</sub> T <sub>1</sub>
15. Providing a judgement or score to describe performance			T <sub>2</sub> T <sub>1</sub> T <sub>3</sub>		

## Teachers' knowledge of assessment

Teachers' knowledge of assessment practices is presented in Table 6.14. The least knowledge is represented by 1 while 4 represents adequate knowledge. Teachers' knowledge of assessment was modest. Although teachers discussed the roles of each party in assessment, modest action was taken to gauge students' readiness for assessment. Assessment was conducted without students' consent for readiness.

Table 6.14: *Knowledge of assessment displayed by individual teacher*

Activity	1	2	3	4
Providing opportunity for students to be assessed when ready		T <sub>2</sub> T <sub>1</sub>	T <sub>3</sub>	
Assessing processes				T <sub>3</sub> T <sub>2</sub> T <sub>1</sub>
Forming groups during practicals			T <sub>2</sub>	
Assessing individuals in group work			T <sub>2</sub>	
Assessing all students in a class on the same skill in one day	T <sub>3</sub> T <sub>2</sub>			
	T <sub>1</sub>			
Perusing students' records		T <sub>3</sub> T <sub>2</sub>		
		T <sub>1</sub>		

1 = little knowledge, 5 = adequate knowledge

All teachers had insufficient knowledge on assessing specific skills instead of all, to an extent that teacher T<sub>1</sub> respondent that the intervention helped very little in assisting them to assess all students in a class at the same time. Knowledge on what exactly to look for when perusing students' records was inadequate. Students did not have proper books for record keeping, but made records on the standard task and assessment materials that they had been given. All teachers assessed students as they were working. As indicated above, only teacher 2 formed groups as more students were involved, but was however not grounded in individual assessment of students in group work.

Teachers appreciated the idea of students assessing themselves and assessment conducted by at least two teachers though they did not practice that. In addition to motivating students, it helped teachers to manage their classes better, particularly in taking care of tools and

implements. Teachers felt that tasks had content that was excessive for both the students and the teachers.

### Resources availability

Resource availability is paramount to the successful implementation of the assessment programme. The provision of resources to schools to enable effective conduct of practicals was a challenge. Due to large class sizes, two to three students shared tools and in some cases up to five shared the same equipment, a situation that hindered effective connection between materials taught and students' experience in the field setting (Finn et al., 2003; Jones, 2006). The situation was serious when more than one class were involved in practicals simultaneously. There were no protective clothing provided for either students or teachers, despite the demand by curriculum for students to conduct practicals, some of which involved the use of chemicals. However, the provision of space in the garden was enough for students to have individual plots. Table 6.15 presents the extent of availability of resources in schools.

Table 6.15: *Availability of physical resources in schools to facilitate performance assessment*

Resources	Availability				
	1	2	3	4	5
Tools			✓		
Equipment	✓				
Other materials	✓				
Garden space					✓
workload		✓			
1= least available	5= most available				

### 6.10.2 Standardising marking

Interviews with teachers revealed high appreciation of the intervention to standardise performance assessment, because the guide they had been using was very subjective resulting in varied interpretations across schools. The assessment syllabus states the number of performance assessments to be made but it did not categorically dictate the level of difficulty. This culminated in schools administering tasks of different demands, format and frequency. The format of performance tasks administered ranged from products assessment, to



interviewing students about their conduct of practicals, to administering written practical tests.

### **6.10.3 Students' understanding of standardised assessment materials**

The outcome of the students' questionnaire is presented in Table 6.16 (below). Three items which were negatively worded were reversed before analysis was done, while two items were removed which correlated very little with the item total. The resulting scale had high internal consistency of 0.80.

Generally, students understood and appreciated the standardised assessment materials as reflected by their high endorsement on most of the statements. However, concern was the 31.6% and 21.1% of students who did not understand the importance of *marking the processes*, and *marking of the products* respectively. This could imply that some students aimlessly conducted practicals, as a result of teachers failing to state the objectives explicitly as discussed above under *instructional approach*. Understanding an individual's role in group work was a problem to 33.3 % of the students.

Also of concern is about a quarter of students (26.3%) who had some difficulties linking the practicals to the theory. One student said: *"In practical, I can understand much rather than theory because one can see how it is done"*. Students liked the idea of being given the assessment materials before the commencement of assessment because it helped them know expectations in advance (Black & Wiliam, 1998; Harlen, 2006), practice first before teacher assessment, and helped them to reconcile the marks with the teacher after scoring: *the teacher marked you according to what you do ...you are given marks you deserve"*.

Students like the idea of being scored by two or more than one teacher:

*"one teacher is not good because s/he can make a mistake during the assessment unlike when they are two ... their scores are going to be different but at some point they would agree that ok these students deserve the mark they have given"*

Table 6.16: *Students' understanding of assessment practices*

<b>The way I have been doing practicals for the past week makes me:</b>	<b>SA</b>	<b>A</b>	<b>D</b>	<b>SD</b>
Understand the importance of practicals	26.3%	68.4%	5.3%	0.0%
Understand the link between practicals and theory	36.8%	36.8%	26.3%	0.0%
Enjoy doing practicals (reversed)	36.8%	52.6%	10.5%	0.0%
Like learning more about the topic	36.8%	52.6%	10.5%	0.0%
Feel encouraged to do practicals (reversed)	47.7%	47.7%	5.3%	0.0%
Understands my role in group work (reversed)	33.3%	33.3%	33.3%	0.0%
Realise the importance of working cooperatively with others	47.4%	42.1%	10.5%	0.0%
Be responsible in caring for tools	47.4%	47.4 %	0.0%	5.3 %
Understand the importance of safety in practicals	31.6%	63.2%	0. %	5.2%
Understand the importance of marking practicals while we are doing them (processes)	15.8%	52.6%	31.6%	0.0%
Understand the importance of marking practicals when we have finished doing them (products).	26.3%	52.6%	21.1%	0.0%

SA = Strongly Agree;      A = Agree;      D = Disagree;      SD = Strongly Disagree

Students also suggested some improvements to be made, such as thorough explanation of assessment materials, and being given a wide choice of what they wanted to do rather than being required to do one thing.

#### **6.10.4 Completion of the assessment instrument (Checklist)**

The completion of the assessment instrument was a problem, with teachers not following the example given in the assessment guide despite prior training. The completion of the form was insufficient and unsystematic. Teachers selected skills to score, for example the form in Table 6.17 (below) shows that the teacher selected skills 1, 3 and 5 but the mark allocated to a certain skill was a total rather than for individual criterion, hence it was extremely difficult to know which criterion was not achieved. Furthermore, not all students were assessed on skill 3 – *calibrating the sprayer*, and no comments were made as to why the students did not achieve a certain criterion. The completion of assessment form revealed their lack of understanding of analytic scoring, resulting in insufficient and unsystematic scoring (Airasian & Russell,

2008). They were used to holistic scoring, as a consequence, the outcome could not be used as the basis for formulating remedial or enrichment strategies (Salvia & Ysseldyke, 1998).

Table 6.17: An example of scoring by teachers

Student Name	Skill						Comments
	Identifying weeds (4)	Organise materials (3)	Calibrating the sprayer (10)	Preparing and spraying chemical (7)	Returning tools & materials to storeroom (4)	Recording transactions (7)	
Lorato Menwana	a (i) (ii) (0) b (i) (ii) (0)	a b c (1)	a b c d (2) e 0 1 2 3 (2) f 0 1 2 3 (3)	a b 0 1 2 3 c 0 1 2 3	a b c d (4)	a 0 1 b 0 1 2 3 c 0 1 2 3	
Lingani Mesho	a (i) (ii) (0) b (i) (ii)	a b c (2)	a b c d (1) e 0 1 2 3 (3) f 0 1 2 3 (3)	a b 0 1 2 3 c 0 1 2 3	a b c d (4)	a 0 1 b 0 1 2 3 c 0 1 2 3	
Tate Mgadla	a (i) (ii) (0) b (i) (ii) (1)	a b c (0)	a b c d (2) e 0 1 2 3 (3) f 0 1 2 3 (1)	a b 0 1 2 3 c 0 1 2 3	a b c d (4)	a 0 1 b 0 1 2 3 c 0 1 2 3	
Ernest Forbes	a (i) (ii) (1) b (i) (ii) (0)	a b c (0)	a b c d (2) e 0 1 2 (0) f 0 1 2 (3)	a b 0 1 2 3 c 0 1 2 3	a b c d (4)	a 0 1 b 0 1 2 3 c 0 1 2 3	
Xenicxi Xhabagkh	a (i) (ii) (1) b (i) (ii) (1)	a b c (0)	a b c d (2) e 0 1 2 3 (0) f 0 1 2 3 (3)	a b 0 1 2 3 c 0 1 2 3	a b c d (4)	a 0 1 b 0 1 2 3 c 0 1 2 3	

### 6.10.5 Record keeping

Record keeping is an important activity during performance assessment, as they are kept to help devise methods that can improve the learner's development (Le Grange & Reddy, 1998). Students did not keep proper records of their practical transactions. They wrote their record on the assessment document they had been given during the practicals due to the conspicuous absence of record books, which is an important material for practicals. Some presented their record in a tabular format, with number of columns and headings differing from one student to the other. Samples of the records kept by students are presented below.

#### Sample 1

Date	Activity	Tools/materials	Reasons
12/08/2010	weeding	Digging fork	To prevent competition for nutrients and water
12/08/2010	watering	Watering can	To encourage the process of photosynthesis To cool the plants

#### Sample 2

Date: 12/08/10

Activity	Reasons	Tools
weeding	To avoid competition of food	Spade
watering	For easy absorption of nutrients in the soil by plants	Watering can
cultivation	To prevent water logging. To encourage good aeration	Digging fork

Others presented their records in a vertical format, placing headings below the other and writing continuously.

### Sample 3

*Date: 12-08-10*

*Activity: weeding*

*Tools: spade, rake*

*Reasons for weeding: to prevent/reduce competition between the weed and the crop for minerals, water, space*

*To maintain tidiness of the plot*

*Reasons spade: to cut/remove the weed with the roots in order for it to not grow again*

*Reasons rake: to obtain a fine tilth after weeding; to obtain tidiness around the plot*

Close scrutiny of the records revealed that reasons advanced were more textbook-like than experienced during practicals. For example, the reasons given for watering in sample 1: “*To encourage the process of photosynthesis*” was more theoretical than practical. As a consequence, record keeping did not reflect any critical thinking by the students. This prompted the researcher to develop a standard form for recording transactions of the activities.

## 6.11 CONCLUSION

Generally, teachers’ instructional practices in assessment were student-centred but the processes of assessment in a student-centred learning approach were found to be insufficient. On the other hand, teachers’ knowledge of assessment was found to be average, possibly due to inadequate training on assessment as initially discussed in Subsection 5.2.3. Teachers therefore needed further training in assessment practices to improve students’ performance.

Resources were found to be insufficient with the exception of the garden space. Equipments and other materials such as protective clothing were found to be in acute shortage. Scoring by teachers was holistic which provided an overview of the students’ performance rather than providing a separate score for each criterion. Teachers found analytic scoring to be presenting more work in classes with more students’ hence high workload. Students were not provided

with record books to record their daily activities, and recording of their transactions was not standardised. However, both teacher and students appreciated the new approach to assessment.

## 6.12 IMPLICATIONS FOR THE SUBSEQUENT DESIGN

The findings implied infusing and modifying a number of factors in the design of the subsequent prototype. Instructions were improved for the intervention to be self-explanatory since the document was given to students in advance, so that they understood it with little explanation from their teachers. Training of teachers was strengthened to include mock-assessment before the actual implementation to ensure that teachers grasped fundamental principles of analytic scoring. Training concentrated on how to select skills to assess a given number of students; how to complete the assessment form; and how to keep record of activities taking place and of assessment outcomes. In addition, a training manual was produced detailing procedure for administering the standard tasks and assessment materials.

A record keeping guide was developed to standardise record keeping throughout the country. Students were guided on how to record activities they carried out. A strategy was devised to encourage them to answer open-ended questions of the questionnaire to provide valuable information to improve the validity of the standard task and assessment materials.

Prior to commencement of lesson observation, the researcher worked with teachers to ensure that they thoroughly prepared for the lesson, such as organising all requisite materials, giving the assessment instrument in advance, explaining the instrument to the students, and how to implement of the intervention. The strategy did not yield positive results and had to be changed.

## CHAPTER SEVEN

### DESIGN, DEVELOPMENT AND EVALUATION OF THE THIRD AND FOURTH PROTOTYPES

#### 7.1 INTRODUCTION

In this chapter, the design of the third prototype is outlined (Section 7.2), and the evaluation of design of the try-out (Section 7.3). Section 7.4 presents the findings of the try-out, and Section 7.5 outlines the characteristics of a practical quality assurance system. The chapter is concluded in Section 7.6.

Design of the third prototype is based on the outcomes of the evaluation of the second prototype, which revealed that the intervention needed improvement in the following areas: instructional practice; the use of summary marksheet; clarity of instructions; the development of the record booklet; guiding teachers during preparation of the lesson; the provision of resources; and innovation change.

#### 7.2 DESIGN OF THE THIRD PROTOTYPE

Design of the third prototype was based on the outcomes of the evaluation of the second prototype, which revealed the following areas that needed reviewing and strengthening:

*Improvement in instructional practice:* teachers' understanding of the use of the tasks and assessment instruments was still unsatisfactory, hence more emphasis was placed on teachers' instructional practices, such as objective of the lesson; advance preparation; teachers' and students' roles during the conduct of the tasks; emphasis on critical thinking; and how to assess.

*Use of the field summary marksheet:* the use of the summary marksheet in the field proved problematic, especially the use of the checklist. This resulted in the modified version presented in Table 7.1 (below). A brief description of each criterion was included in this version to facilitate quick remembrance of each criterion during assessment.



*Clarity of instructions:* further improvements were made on the instructions as the document was given to students to study in advance. They needed to understand it when reading it on their own. Improvements concentrated on how to select skills to assess a given number of students; how to complete the assessment form; how students should keep record of activities; and record-keeping of assessment outcomes by teachers.

*The development of record keeping booklet:* a standard record-keeping format was developed, aligned to the assessment instrument. It guided students on how to keep record of activities carried out during the conduct of practicals (See Appendix 4.5).

*Guiding teachers during the preparation for the lesson:* before the observation of the lesson commenced, the researcher worked with the teachers to prepare thoroughly for the lesson. The researcher guided teachers on what to do and how to do it. These preparations included organising all materials needed; how to ask divergent questions which are thought-provoking; how to link the practical to everyday life experiences; how to state the objective of the lesson; and how to assess.

*Provision of resources:* some resources were available in schools, but were not optimally used. Emphasis was placed on optimal usage of available resources, given that resources for performance assessment were costly to provide for. For example, garden space was abundant in almost all schools, even though teachers made students share plots.

*Helping teachers to embrace change:* although teachers embraced the intervention, they needed something that could easily be used. They considered tasks and assessment to be placing too much demands on students and suggested lowering the level. However, tasks were maintained as they were, since they required students to think critically to construct their own solutions. Rather, emphasis was placed on teachers accepting the paradigm shift in assessment.

Table 7.1: Example of summary marksheet with brief notes for each criterion for field work

**Total: 29 marks**

Student Name	Skill					Total Marks	Comment
	7. Determining the fertiliser requirements (2)	8. Selecting tools and fertiliser (3)	9. Weighing the fertiliser (5)	10. Applying fertiliser (5)	11. Returning tools & materials to storeroom (4)		
Lorato Menwana	a ) Determine the soil Ph b ) What & when to apply fertilisers	a) Identify tools for application b ) protective clothing c) Remove fertiliser to place of weighing	a) Zeroing the scale b) Reading of container alone c ) reading of fert + container ...(1) d) Fertiliser + container reading (2) e) Reading (2)-(1)	a) Use correct method b) Use tools correctly, c) Apply to the correct depth, d) Avoid fertiliser-plant contact e) Avoid skin contact.	d) Clean all tools e) Carry tools & materials safely f) Place tools & materials properly. g) Work diligently with minimal supervision		
Lingani Mesho	a ) Determine the soil pH b) What & when to apply fertilisers	a) Identify tools for application b) protective clothing c ) Remove fertiliser to place of weighing	a ) Zeroing the scale b) Reading of container alone c ) Reading of fert + container (1) d) Fertiliser + container reading (2) e ) Reading (2)-(1)	a) Use correct method b) Use tools correctly, c) Apply to the correct depth, d) Avoid fertiliser-plant contact e) Avoid skin contact.	a) Clean all tools b) Carry tools &materials safely c)Place tools &materials properly d) Work diligently with minimal supervision		
Tate Mgadla	a) Determine the soil Ph b) what & when to apply fertilisers	a) Identify tools for application b) protective clothing c) Remove fertiliser to place of weighing	a) Zeroing the scale b)Reading of container alone c ) Reading of fert + container (1) d) Fertiliser + container reading (2) e) Reading (2)-(1)	a) Use correct method b) Use tools correctly, c ) Apply to the correct depth, d) Avoid fertiliser-plant contact e) Avoid skin contact.	a) Clean all tools b) Carry tools &materials safely c)Place tools &materials properly. d) Work diligently with minimal supervision		
Ernest Forbes	a) Determine the soil Ph b) what & when to apply fertilisers	a) Identify tools for application b) protective clothing c) Remove fertiliser to place of weighing	a) Zeroing the scale b) Reading of container alone c ) Reading of fert + container (1) d) Fertiliser + container reading (2) e) Reading (2)-(1)	a) Use correct method b) Use tools correctly, c) Apply to the correct depth, d) Avoid fertiliser-plant contact e) Avoid skin contact.	a) Clean all tools b) Carry tools &materials safely c) Place tools &materials properly d) Work diligently with minimal supervision		
Nonny Meshack	a) Determine the soil Ph b) what & when to apply fertilisers	a) Identify tools for application b) protective clothing c)Remove fertiliser to place of weighing	a) Zeroing the scale b) Reading of container alone c) Reading of fert + container (1) d) Fertiliser + container reading (2) e) Reading (2)-(1)	a) Use correct method b) Use tools correctly, c) Apply to the correct depth, d) Avoid fertiliser-plant contact e) Avoid skin contact.	a) Clean all tools b) Carry tools &materials safely c) Place tools &materials properly d) Work diligently with minimal supervision		

Teacher's Name \_\_\_\_\_

Teacher's Signature \_\_\_\_\_

Date \_\_\_\_\_

Snr Teacher's Name \_\_\_\_\_

Snr Teacher's Signature \_\_\_\_\_

Date \_\_\_\_\_

### 7.3 EVALUATION DESIGN OF THE TRY OUT

The evaluation of the third prototype was aimed at determining the expected practicality of the exemplar assessment materials for agriculture at Form Four level. The evaluation was through observation of students and teachers during the conduct of the performance assessment, teachers and students completing questionnaires, and teaches and students interviews. The questionnaires and interviews complemented the observation. Eight of the initial nine teachers were able to implement the intervention.

#### **7.3.1 Aim and research question**

The third prototype was tried out to evaluate the criterion of practicality as discussed in Subsection 4.5.1. Practicality was determined by means of the standard tasks and assessment materials' ability to meet the criteria of timeliness, cost of implementation, utility, and ease of understanding when assessing Form Four Agriculture students' practicals. The evaluation was guided by the question:

*How can quality assurance processes for performance assessment be developed to ensure valid and reliable marks?*

The development of a quality processes was a continuous cyclic iterative process involving stakeholders at different levels of the development. The prototype was an improvement of the first two prototypes. The ultimate goal was to implement the final prototype in the real field situation.

#### **7.3.2 Research design**

Eight teachers from three schools were involved in the in-depth study of how practicable the tasks and assessment materials were. Data collection was triangulated by the use of different sources (teachers and students) using different instruments, such as observation schedule, teacher questionnaire, student questionnaire, teacher interview and student interview, to enhance corroboration of findings (Creswell & Miller, 2000; Mertens, 2010; Patton, 2002).

### 7.3.3 *Participants*

Teachers and students participating in this study were drawn from different schools that offered Agriculture. The classes of students selected were by virtue of their teacher's participation.

#### *Schools*

A total of three government schools from two regions in proximity to the researcher were involved in the study. The sample was small because when working within the theory of constructivism, the goal is to identify information-rich cases that will allow studying a case in depth (Mertens, 2010). The criteria for purposively sampling were: at least one school in rural and one in urban centre; proximity to the researcher; administrators' willingness to support the study; and teachers' willingness to participate. However, it must be noted that Botswana government schools are standard in terms of resources allocation, staffing, enrolment, and work planning (Motswiri, 2004; Yandilla et al., 2003).

#### *Teachers*

Nine teachers and their students from three schools were targeted for participation in this study. Due to changes in the timetable, teacher T<sub>8</sub> withdrew from participation because the changes were not convenient for both of us. Three teachers were purposively sampled (Cohen, Manion & Morrison, 2000) from school **A** and **C**, while only two were sampled from school **B**, to implement the intervention. Purposive sampling was used to select willing teachers to advance insight into the classroom assessment dynamics, as teachers were found to be de-motivated by having to implement performance assessment (Keightley & Coleman, 2002). Background information for teachers involved in the study is presented in Table 7.2 (below).

Table 7.2: *Background information of respondents*

Variable	School A				School C		School B	
	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	T <sub>7</sub>	T <sub>9</sub>
Task	1	1	1	3	3	3	2	2
Age	43	39	40-45	37	40	34	37	39
Sex	M	F	M	F	M	F	M	F
Academic Qualification	BSc AgricEd	BSC AgricEd	BSc	BSC AgricEd	BSC AgricEd	BSc	BSc	BSC AgricEd
Prof Qualification	BSc AgricEd	BSC AgricEd	PGDE	BSC AgricEd	BSC AgricEd	PGDE	PGDE	BSC AgricEd
TE(yrs)	15	16	16	14	16	8	8.5	15
Class size	26	24	33	35	35	35	35	31
TESS (yrs)	2	7	16	4	4	7	8	5

TE = Teaching Experience TESS = Teaching Experience in Senior School

### *Students*

Students involved in the study were those from the classes of participating teachers. Since each teacher had more than one class, purposive sampling of one class for each of the eight teachers was carried out (McDaniel & Gates, 2010). A total of 254 students were involved in completing the questionnaire, 177 returned completed questionnaires (69.7%) of whom 80 were boys, 96 girls, and one did not indicate sex. Table 7.2 (above) shows the total number of students in the class selected for the study. Students were observed during the conduct of performance assessment and later completed a questionnaire. Group interviews of six students per task were conducted to obtain their views on the experience of implementing the standard tasks.

### **7.3.4 Data Collection Strategies**

Data collection was triangulated to establish convergence of evidence among multiple varied sources of data and methods in an effort to overcome the inherent weaknesses of each, and to

minimise uncertainty in data interpretation (Creswell & Miller, 2000; Patton, 2002). Teachers and students were observed during the conduct of the performance assessment, completed a questionnaire and then interviewed. The data collection methods are described below.

#### *Lesson observation*

All eight teachers who participated in the study were observed conducting performance assessment. According to Fink (2005), observations are appropriate for obtaining global portraits of the dynamics of a situation. Teachers conducted the performance assessment in one week, which facilitated observation of the lessons by the researcher. An observation schedule (Appendix 4.5), described in Subsection 4.5.3 was used to collect data during the activities of the lesson. During the proceedings, the researcher posed as a complete observer (Mertens, 2010).

#### *Teacher questionnaire*

At the end of the lesson, the teacher completed a questionnaire which sought his or her views on the practicality of the standard tasks and assessment materials. The questionnaire had both close-ended and open-ended questions. Close-ended questions targeted teachers' instructional behaviour, knowledge of assessment, standardising of assessment and class-management. Open-ended questions sought the views of teachers on quality of the task, content of task, format of task and language used on the tasks and assessment materials.

#### *Teacher interview schedule*

A semi-structured interview (Forrester, 2010; Mertens, 2010) was administered at the end of the lesson, the aim was to capture respondents' views about the impression of the intervention. Issues discussed ranged from the usefulness of the standard tasks and assessment materials, their feasibility, things they did not like, things they liked, and how the assessment could be improved.

#### *Student interview*

A focus group interview of six students per task was conducted at the end of the last day of lesson observations. Each teacher had to conveniently select two students to form a group of six. The interview schedule was semi-structured, consisting of nine questions which were posed in the same way from one group to the other. Probing and follow-up questions were

intended to elicit more information (McIntire & Miller, 2007), to get insight into students' views about the standard task and assessment materials they had been implementing. The interview lasted for about fifty minutes and transactions were audio-taped for later transcription.

### *Student questionnaire*

Students completed a questionnaire at the end of implementing the intervention. Teachers distributed questionnaires to their students and later collected them on behalf of the researcher. The questionnaire sought to find students' opinions on the intervention. The questionnaire consisted of (i) a scale and (ii) open-ended questions.

### **7.3.5 Procedure**

Teachers who participated in the implementation of the intervention were subjected to rigorous training to facilitate easy and uniform understanding. Training was carried out a few weeks before the implementation so as to allow ample time for conceptualisation. Teachers were supplied with all materials needed during training, such as the task, teacher's guide, summary marksheet, assessment instrument, and student's recordkeeping booklet. Training which lasted for 2-3 hours was conducted at the respective schools in the afternoon when lessons were over.

The implementation of the tasks followed each other sequentially. Each of the three schools implemented one task. Task 1 was implemented first at school **A**, followed by task 2 at school **C** the following week, while task 3 was done some weeks later at school **B**. All the three teachers in the same school started their tasks during the same week, and in most cases tasks were completed within one to two days. This facilitated visits to schools by the researcher.

During the day of the observation, the researcher arrived well in time to observe preparatory steps. The lesson started with the teacher explaining the objective of the lesson, clarifying students' and the teacher's roles, outlining expectations from students, and stating how the assessment would be conducted. After a short discussion between the teacher and students, the class left for the garden (site of implementation). Students were observed removing tools from the storeroom and carrying them to their plots. It was difficult to observe the execution

of all the skills for all the three teachers. As a result, observation of one teacher was made at a time, even though they might be working in the garden simultaneously.

For example, in task 1, Teacher  $T_1$  and  $T_2$  started at the same time and students for  $T_1$  were observed the first day, while  $T_2$  was observed the second day. Skills observed for  $T_1$ 's students were 1, 2, 4, and 5, while for  $T_2$ 's were 2, 3, 4, and 5. Thus  $T_1$ 's and  $T_2$ 's students were not observed on skills 3 and 1 respectively. According to skills equating, discussed in Subsection 6.3.2, skills 1 and 3 are equivalent in terms of demand, hence students were not advantaged or disadvantaged by assessing them on different skills. The same explanation applies to tasks 2 and 3. Teacher  $T_3$ 's students were observed implementing the same skills as  $T_1$  at a later date. Tasks assessed by each teacher are shown in Figure 7.1 (below).

During the implementation of each task, the researcher observed and completed an observation schedule. At the end of the lesson, teachers and students completed a self-administered questionnaire to reflect on how they perceived the practicality of the exemplar tasks and assessment materials. Students' record books were perused to check how teachers scored the work, and if they subsequently transferred marks from the summary marksheets to the scoring instrument for individual student (checklist and scale). Students were interviewed at the end of the last lesson observation and appointments made with teachers for interviews later. All the interviews were audio-recorded and transcribed verbatim (see Appendix 7.2).



TASK 1: Preparing a plot and planting

School	Teacher	Skill				
		Preparing a plot (1)	Using tools (2)	Planting (3)	Return tools & materials to s/room (4)	Recording transactions (5)
A	T <sub>1</sub>	█	█		█	█
	T <sub>2</sub>		█	█	█	█
	T <sub>3</sub>	█	█		█	█

TASK 2: Applying fertiliser as basal dressing

School	Teacher	Skill					
		Determining the need to top dress (1)	Selecting tools and fertilisers (2)	Weighing the fertiliser (3)	Applying Fertiliser(4)	Return tools & materials to s/room (5)	Recording transactions (6)
C	T <sub>4</sub>	█	█			█	█
	T <sub>5</sub>	█	█			█	█
	T <sub>6</sub>	█	█		█	█	█

TASK 3: Controlling weeds using chemicals

School	Teacher	Skill					
		Identifying weeds (1)	Organising materials (2)	Calibrating the sprayer (3)	Preparing and spraying the chemical (4)	Returning tools & materials to s/room (5)	Recording transactions (6)
B	T <sub>7</sub>	█	█	█		█	█
	T <sub>9</sub>	█	█		█	█	█

*Numbers in bracket represent skill number*

*Figure 7.1: Implementation Plan*

## 7.4 FINDINGS OF THE TRY OUT

This section describes the results of the try-out carried out to evaluate the impact of quality assessment materials in agriculture at Form Four level in Botswana schools. The evaluation focussed on how the intervention was implemented, and the results presented were based on participants' experiences with the intervention and lesson observations.

### 7.4.1 *Participants' experiences with the intervention*

Teachers and students implemented the quality standard tasks and assessment materials and subsequently completely questionnaires. They were later interviewed to get their views about the intervention. Essentially, three themes emerged from participants' experiences with the intervention:

#### *a) Overall impression*

Teachers were generally impressed by the implementation of the intervention and rated the assessment instruments to be of high quality compared to what they had been using (See Section 2.9 for elaboration of the instrument previously used). They indicated that the instrument clearly spelled out what needed to be done by the students, and hence were easy for students to follow on their own. Students were impressed by the idea of being given the instrument in advance so that they could study and assess themselves before the teacher did. By making assessment a public domain helped students to know the teacher's expectation and adequately prepared themselves. One student commented: *S504: it helps me to know what I am expected to do as a way of earning marks.* Teacher T6 commenting on the same subject said: *nothing is hidden from them, even the marks are there, so that is one thing that I like about it.* Students also engaged in self-assessment and peer assessment which enhanced their achievement (Black & Wiliam, 1998). Peer assessment facilitated collaboration and cooperation in learning ill-structured problems (Burriss & Garton, 2007), helping them to develop critical thinking skills in the process.

Students' knowledge of the purpose of assessment and access to assessment instrument prior to commencement of teacher assessment also enhanced their morale (Salvia & Ysseldyke, 1998), resulting in commitment to their work, as indicated by both teachers and students: *S320: it gave me an opportunity to be able to work hard every time I did my practicals. S321:*

*It will help me to be a hard worker all the time and also improve my points. Teacher T4 said: they knew what was expected of them and this time there was much improvement.*

Furthermore, understanding the objectives of the performance assessment made students to be mature and responsible for their learning, thus making teachers' work easier. Teacher T6 commented thus: *... in the past, we used to have a student who runs away, without taking tools. Nowadays they know these marks and it's a policy now.* Teachers were earnestly waiting for the instrument to be introduced officially in all the schools for use to generate valid and reliable marks.

*b) Improvement in learning*

Improvement in learning can be viewed in terms of standardising assessment throughout the country, imparting critical thinking skills, holistic assessment of the students, facilitation of feedback, transparency in assessment and students' motivation.

Standardisation of assessment

Although it is difficult to develop congruent tasks, it is important that the tasks be equivalent when reliability is an important quality measure (Lennox, 2000). Previously, schools or even individual teachers designed their own assessment criteria (Section 5.4), based on the criteria outlined in Section 2.9. As a consequence, standards varied between schools and even between teachers within the same school, as the instruments lacked both content and construct validity (Ary et al., 2006; Linn et al, 1991). Some teachers took advantage of these unclear statements to award students marks without conducting any performance tasks.

The resulting outcomes were consequentially invalid (Messick, 1989) and unreliable (Salvia & Ysseldyke, 1998). The developed standard tasks and assessment instrument therefore had the ability to standardise assessment across the country for certification purposes, because of the teachers' observational assessment of the student's skills using clearly defined criteria (Airasian, 2005; Airasian & Russell, 2008; Hargreaves, 2007), thus resulting in more objective assessment.

## Imparting critical thinking skills

Participants' impression of the intervention was that tasks covered broad and in-depth content of knowledge and skills (Diez, 2002; Rennert-Ariev, 2005; Ryan, 2006), contrary to Lane and Stone's (2006) contention that performance tasks cover little content. For example, one student S318, when responding to a question seeking to know what s/he did not like about the assessment strategy, said: *it needed a lot of concentration, and hard work... it made us learn things that we were not aware of.*

Students reported applying a multi-faceted approach to problem-solving which generated varied solutions, as evidenced from scrutiny of their record keeping, leading to improvement in learning (ARG, 2006; Black & William, 1998; Crooks, 2004; McMillan, 2004), and acquisition of both knowledge and skills (Burris & Garton, 2007). This was affirmed by Teacher T7:

*... this time it was good, because I gave them the task well in hand, they read and we selected the skills that would be assessed. Even themselves, students, when they went to the garden, they knew what was expected of them and this time there was much improvement, ...so they were able to do a better job this time.*

Apart from assessing complex thinking skills, performance assessment provided some pupils who did poorly on selection type tests the opportunity to show their achievement in an alternative way (Ryan & Miyasaka, 1995), as affirmed by student S326 who said: *it gives us, students, knowledge on practicals because if one does not understand the theory he/she can focus on the practicals.*

## Holistic Assessment

Contrary to previous practice in Botswana schools, the assessment instrument emphasised processes across all domains of cognitive, psychomotor and affect. Students were engaged in performing tasks which enhanced the application of critical thinking of which the processes were assessed. The product was assessed as it was in some cases the focus of assessment (Gronlund, 2003; Stiggins, 1997; Thorndike & Thorndike-Christ, 2010). Both students and teachers appreciated the objective assessment of products and processes made possible through the use of detailed criteria (Nitko, 2004). The assessment of affective skills was even appreciated by students. Student S111 said: *I like being observed how I handle tools,*

*cooperate with others in terms of language we use on each other, how I do my plot and all my practicals...*

### Feedback

The assessment process provided feedback from the peers, teacher or from self, as students had access to the instrument well in advance of teacher assessment. Teachers gave students feedback on their performance and students agreed that feedback resulted in improved achievement. Improved performance was reported by Christmann and Budgett (2003), Nir-Gal and Klein (2004) and Thomas, Davis and Kazlauskas (2007), who found that children learned more and scored higher on all the cognitive measures of abstract thinking, planning, vocabulary, and reflective thinking when using computers in the presence of mediating adults.

### Students' motivation

In Section 5.4, it was indicated that students' attitudes towards performance assessment was negative, because they never knew why they were assessed in performance tasks. However, after the intervention, students' attitude had changed. The knowledge of why assessment was conducted and what was expected of them had a significant impact on their motivation, as argued by both Weiner (cited in Torrance & Pryor, 1998) and Harlen (2006). Students tended to focus on their learning goals and chose challenging tasks irrespective of their ability with the aim of succeeding (Dweck cited, in Torrance & Pryor, 1998). The strategy of feedback, coupled with the opportunity for reassessment, motivated students to do more than before (Harlen, 2006). Two students, S316 and S108, said the following about their motivation, resulting from the use of the new assessment instrument S316: *It motivates the students, because the teacher gives the student's advice on what is required.* S108: *they have motivated me to be a good and responsible person in agricultural sector.*

Teachers also echoed students' opinions regarding the intervention's impact on motivation. One teacher T6 said: *... it makes the students to enjoy their work and to appreciate, they can even be aware of how much they can get from the practical.* While the other teacher T8 said: *It's a perfect one, because this really made them open and were interested and knew what they were doing.*

## Transparency in assessment

The openness of the tasks and the numerous skills in each task facilitated access by students to different levels of ability. Students had a choice of the task to do as echoed by one student: *S503: you chose your own practical which you prefer*. Before assessing, teachers negotiated with students to determine their status of readiness. Consulting students resulted in better diagnosis of students' strengths and weaknesses to be in a better position to help them improve their learning.

Students indicated that they were given the chance to master the skills before any assessment could be carried out. Student S110 remarked: *A student is given a chance by the teacher to make sure that the practical is in good condition before marking*, while S120 said, *well every student was given an opportunity to showcase how hard working we are, determined, willing to do our school work without being forced*.

### *c) Implementation Hiccups*

Prototype development of the exemplar assessment materials in collaboration with practitioners and experts was meant to identify problems during the development cycles and institute corrective action before rolling out. As the intention of prototyping is to identify hiccups which could hinder effective implementation of the intervention, the following problems were identified: Inadequacy of resources; teachers' resistance to change; task length; and cumbersomeness of the instrument.

Time as finite resource was mentioned as a major problem. Despite disagreement in literature on whether small class sizes result in improved achievement (See Sections 3.5 and 3.8), reduced class sizes in the context of agriculture would result in matching students to both physical and time resources, culminating in reduced teacher workloads and facilitating ease of assessment. A number of teachers expressed doubt with regards to the successful implementation of the intervention in schools, given the current status of resources.

To compel schools to acquire the necessary resources, the assessment instrument outlines all prerequisite resources needed for effective implementation of each task, and requires schools to be accredited to implement the tasks. In terms of the infrastructure, particularly the school garden, which was considered as the laboratory for agriculture practicals, there was enough space which was not utilised optimally. The problem with school garden was its vulnerability

to displacement by other infrastructural developments, despite the requirement by the Revised National Policy on Education of 1994 for schools to have one. Teachers' workloads were high as a result of schools having only one garden assistant, who has no formal training in agriculture.

Teachers lacked understanding of how performance assessment should be done. In large classes of 35 or more students, it was not practical to assess all students. The introduction of skills equating discussed in Sub-section 6.3.2 was meant to circumvent this problem. Teachers need more training on how this is done. Teachers unanimously raised the issue of the cumbersome nature of the assessment instrument, and suggested putting everything in one page to avoid too much paper work. Although putting everything on one page was desirable, it was not practically possible. In addition, it would result in excluding much valuable information, culminating in superficial assessment and completely deviating from the intended purpose of developing an assessment instrument for producing reliable marks.

Transferring marks from the field summary marksheets to the checklist was considered extra work by teachers, compounding their already overloaded schedules. There is no doubt that the assessment strategy would result in huge paperwork (Brown, 1999; Collins, 1999; Collins et al., 2004) which would need handling and ample storage space. However, training to change teachers' mind set to embrace assessment for learning instead of assessment of learning is viewed as the ultimate solution to the problem of performance assessment in agriculture.

#### **7.4.2 Lesson Observations**

The lesson observations provided the researcher the opportunity to collect firsthand information on students' activities during the conduct of performance assessment. Lesson observations were based on instructional behaviour, knowledge of assessment, record-keeping by students, and scoring of students' work.

##### **Instructional behaviour**

Teachers instructional behaviour generally improved compared to the time of piloting. Table 7.3 (below) presents the results of teachers' instructional activities. It was observed that teachers' instructional activities were geared towards assessment for learning, which supports and improves students' learning and motivation (ARG, 2002; Crooks, 2004; Taylor, 2004).

The only instructional activity which was moderately emphasised was *reasoning skills*, most likely due to teachers' insufficient knowledge of the proposed change (Ertmer, 1999) and personal factors that are ingrained, such as instructor's beliefs about the instructional process and the value the change brings (Harrington, McElroy & Morrow, 1990; Kent & McNergney, 1999). Generally, teacher T<sub>2</sub> exhibited low understanding for assessment of learning.



Table 7.3: *The extent of teachers' embraced assessment for learning*

<b>Instructional activities</b>		<b>To some extent</b>	<b>To a moderate extent</b>	<b>To a great extent</b>	<b>To a very great extent</b>
1.	The teacher distributes the task to the class before the start of the practicals				T <sub>5</sub> T <sub>7</sub> T <sub>6</sub> T <sub>1</sub> T <sub>2</sub> T <sub>3</sub> T <sub>4</sub> T <sub>8</sub>
2.	The teacher states the performance assessment objective before the start of the practicals			T <sub>6</sub> T <sub>1</sub> T <sub>2</sub> T <sub>3</sub> T <sub>4</sub>	T <sub>5</sub> T <sub>8</sub> T <sub>7</sub>
3.	The teacher asks questions to gauge the students level of knowledge of the activity		T <sub>4</sub>	T <sub>2</sub> T <sub>1</sub> T <sub>3</sub> T <sub>5</sub> T <sub>7</sub> T <sub>6</sub> T <sub>8</sub>	
4.	The teacher links the relevance of the practical to everyday lives		T <sub>4</sub> T <sub>2</sub>	T <sub>1</sub> T <sub>3</sub> T <sub>8</sub> T <sub>7</sub> T <sub>6</sub>	T <sub>5</sub>
5.	The teacher clarifies what resources are to be used and how they are to be used			T <sub>1</sub> T <sub>6</sub> T <sub>2</sub> T <sub>4</sub>	T <sub>5</sub> T <sub>7</sub> T <sub>3</sub> T <sub>8</sub>
6.	The teacher spells out observable aspects of the student's performance/product that should be judged	T <sub>2</sub>		T <sub>1</sub> T <sub>3</sub> T <sub>8</sub> T <sub>7</sub> T <sub>4</sub>	T <sub>5</sub> T <sub>6</sub>
7.	Teacher clarifies what will be assessed and how	T <sub>2</sub>		T <sub>1</sub> T <sub>3</sub> T <sub>8</sub>	T <sub>5</sub> T <sub>7</sub> T <sub>6</sub> T <sub>4</sub>
8.	The teacher stresses observation of safety	T <sub>2</sub>	T <sub>1</sub>	T <sub>4</sub> T <sub>5</sub> T <sub>8</sub> T <sub>6</sub>	T <sub>7</sub> T <sub>3</sub>
9.	The teacher states the students' task			T <sub>2</sub> T <sub>3</sub> T <sub>6</sub>	T <sub>5</sub> T <sub>7</sub> T <sub>1</sub> T <sub>4</sub> T <sub>8</sub>
10.	The teacher states the teacher's role	T <sub>2</sub>		T <sub>6</sub> T <sub>3</sub> T <sub>4</sub>	T <sub>5</sub> T <sub>7</sub> T <sub>1</sub> T <sub>8</sub>

11. The teachers emphasises reasoning as opposed to rote learning	T <sub>3</sub>	T <sub>5</sub> T <sub>6</sub> T <sub>2</sub> T <sub>4</sub>	T <sub>8</sub> T <sub>7</sub>	T <sub>1</sub>
12. The teacher organises material for the practical			T <sub>1</sub> T <sub>3</sub> T <sub>4</sub> T <sub>5</sub> T <sub>7</sub> T <sub>6</sub> T <sub>8</sub>	T <sub>2</sub>
13. The teacher manages time well		T <sub>3</sub>	T <sub>1</sub> T <sub>5</sub> T <sub>7</sub> T <sub>6</sub> T <sub>4</sub> T <sub>8</sub>	T <sub>2</sub>
14. The teacher provides an appropriate setting to elicit and judge the performance or product			T <sub>2</sub> T <sub>1</sub> T <sub>4</sub> T <sub>6</sub>	T <sub>3</sub> T <sub>5</sub> T <sub>8</sub> T <sub>7</sub>
15. The teacher provides a judgement or score to describe performance		T <sub>1</sub>	T <sub>2</sub> T <sub>4</sub> T <sub>5</sub> T <sub>8</sub> T <sub>7</sub> T <sub>6</sub>	T <sub>3</sub>

## Knowledge of assessment

Teachers' knowledge of assessment practices is presented in Table 7.4 (below). The least knowledge is represented by 1, while 4 represents adequate knowledge of assessment. Teachers are represented by T<sub>1</sub> to T<sub>8</sub>. It was observed that teachers' knowledge of assessment was above average to adequate in most cases, which enabled them to practice assessment for learning.

Table 7.4: *The extent of teachers' knowledge of assessment*

Variable	Teachers' extent of knowledge			
	1	2	3	4
The teacher provides opportunity for students to be assessed when ready	T <sub>3</sub>		T <sub>5</sub> T <sub>2</sub> T <sub>4</sub> T <sub>3</sub> T <sub>7</sub>	T <sub>1</sub> T <sub>8</sub>
The teacher assesses processes			T <sub>5</sub> T <sub>2</sub> T <sub>4</sub> T <sub>3</sub> T <sub>7</sub>	T <sub>3</sub> T <sub>1</sub> T <sub>8</sub>
The teacher forms groups during practicals			T <sub>5</sub> T <sub>8</sub> T <sub>7</sub> T <sub>4</sub>	T <sub>3</sub> T <sub>2</sub> T <sub>1</sub> T <sub>3</sub>
The teacher assesses individuals in group work	T <sub>5</sub> T <sub>1</sub> T <sub>7</sub>			T <sub>3</sub> T <sub>2</sub> T <sub>8</sub> T <sub>4</sub> T <sub>3</sub>
The teacher assesses all students in a class on the same skill in one day	T <sub>1</sub> T <sub>4</sub> T <sub>3</sub>		T <sub>5</sub> T <sub>2</sub> T <sub>8</sub>	T <sub>3</sub>
The teacher peruses students' records	T <sub>8</sub>	T <sub>3</sub> T <sub>5</sub> T <sub>2</sub>	T <sub>1</sub>	
		T <sub>4</sub> T <sub>3</sub> T <sub>7</sub>		

However, some few past experiences were deeply entrenched in teachers' practices as observed by Harrington, McElroy and Morrow (1990) and Kent and McNergney (1999), such as the desire to assess all students in a particular skill at the same time, and continuous marking of students' record books for feedback purposes. This is evidenced by teacher T6: *Why can't we ... go at once and mark for the whole class.*

## Students' record-keeping

Students' record keeping had improved as students kept detailed records that reflected practical transactions. The development of a guide on record-keeping helped to standardise keeping of records. Students also understood the importance of keeping records, as evidenced by the following quote when students were asked about things they liked about this way of marking practicals: *S308: writing record in record sheet then they are marked.* Teachers also

were impressed with students' understanding of the importance of keeping records. Teacher T8 said:

*... the other time I was not there, then I said, "You go to the garden with the other teacher, when I come back I'm going to see the records." When I came back I could see that they had recorded everything that they did with the other teacher.*

Despite good records kept by students, teachers did not peruse them to provide students with necessary feedback.

Scoring of students work

Assessment data is often used to inform appropriate instructional strategies (Thorndike & Thorndike-Christ, 2010). When assessment proceeds in a haphazard manner, the information collected is neither reliable nor valid, and misdirects instruction. The completion of the assessment instrument improved after modifying the field marksheet by including brief description of each criterion to aid teachers during scoring. However, the completion was not thorough, an indication that further improvement and training of teachers were needed before making important decisions concerning students based on its use (Stiggins, 1997).

## 7.5 CHARACTERISTICS OF A PRACTICAL QUALITY ASSURANCE SYSTEM

The sub-research question (e) of the second main research question 2 (Section 1.5) set out to determine the characteristics of an effective quality assurance system for ensuring valid and reliable performance assessment in agriculture in Botswana. However, since the intervention could not be field tested, only the characteristics of the practicality of the intervention could be inferred (see Table 4.3, above).

**Assessment policy:** Policy formulation is the first and foremost quality assurance aspect to be satisfied. The policy guides the schools on how to conduct performance assessment, who should conduct performance assessment, how many tasks should be done, what is the role of the student, what resources schools should provide, how the marks should be stored, and what supervision is needed.

**Trained teachers:** the findings from a baseline survey indicated that teachers training to conduct performance assessment was lacking. However, prototype development of tasks and assessment materials revealed that if teachers are well trained to conduct performance assessment, valid and reliable outcomes are achieved which improve students learning. Training teachers to equip them with the necessary performance assessment skills is therefore an important characteristic of an effective performance assessment system. Once teachers have been trained, it is imperative that they are accredited to conduct performance assessment and that the accreditation be renewed periodically.

**An efficient monitoring system:** monitoring of school-based performance assessment should be thorough, starting with the teachers' immediate supervisor ascending to the school administrators and finally Ministry officials. For supervisors to execute this mandate effectively, they need training as well on the conduct of performance assessment. Training programmes targeting supervisors should be developed, with emphasis on classroom-based assessment.

**Availability of student-centred standard tasks and assessment materials:** it is important to provide exemplar materials that have been iteratively developed in collaboration with practitioners and other stakeholders. The exemplar assessment materials should comprise the task and assessment instrument accompanied by the administration manual detailing how the task should be administered and assessment.

The tasks should have the following characteristics: i) physical accessibility to all students: ii) approachable in multiple ways iii) inclusion of group activities iv) provision of opportunities for all students to interact and cooperate within a group v) accommodation of special needs students vi) incorporation of high order thinking activities vii) clearly written instructions and viii) activities targeting different levels of ability.

The assessment instrument should consist of: i) detailed criteria with clearly written instructions ii) assessment of processes, product and affect domains iii) provision for multiple assessments and iv) provision for validation of assessment.

**Sufficient provision of resources:** Performance assessment requires individualised assessment, which is very time-consuming. For teachers to effectively conduct performance assessment for certification, they need to have sufficient resources. Resources needed for performance assessment in Agriculture are standard tasks and assessment materials, time,

multiple assessors, tools and equipment. The present teachers' workloads can be reduced if adequate resources are not provided, because this would result in the reduced student/teacher ratio to afford more contact time.

**Multiple modes of assessment:** A more valid and reliable assessment using different methods such as observation (processes), product, and affect across a range of situations, (Airasian, 2005; Mamary, 2007) produced different types of data reflecting different achievements (Tindall & Marston, 1990; Stiggins, 1997). Students' feelings, values, attitudes and emotions have to be constantly assessed to guide appropriate remedial work.

**Multiple rating:** when using multiple observations of students' performance more reliable and accurate information (Airasian, 2005) was produced which was more acceptable to both parties. Rudner (1994) asserts that multiple raters can improve reliability just as multiple test items can improve the reliability of standardised tests, if it is done clearly crafted criteria and quality assurance procedures put in place.

## 7.6 CONCLUSION

The implementation of the exemplar assessment material was well received by both teachers and students, resulting in improved outcomes. When students fully understood why and how they were assessed, they were motivated and tended to take responsibility for their assessment to work hard to achieve goals they set for themselves. Students greatly appreciated the idea of being given the instrument in advance to study and assess themselves prior to the teacher assessment. The transparency of the assessment allowed students the opportunity to interact with other students, consequently learning from each other in a collaborative environment. Such collaboration resulted in the acquisition of very important life skills and development of abstract thinking. Apart from imparting critical thinking skills, performance assessment provided some pupils with an alternative way to prove their ability in a different way.

This way of assessing was found to be objective, and motivated both teachers and students because the same standard was being used throughout. Consequently, class management improved, resulting in reduced work for the teacher. Teacher assessment practices also improved, indicating understanding of the conduct of performance assessment. Thus, the

characteristics of an effective quality assurance system for ensuring valid and reliable performance assessment were enumerated as development of student-centred exemplar assessment materials, accreditation of teachers, a strong monitoring system, approval of schools to implement performance assessment, provision of resources, multiple modes of assessment using multiple raters.

## CHAPTER EIGHT

### SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

#### 8.1 INTRODUCTION

This chapter outlines the conclusions reached in this study which sought to find out the validity and reliability of performance assessment practices in Botswana, and then develop quality assurance processes to improve it. It starts by presenting the summary of the study in Section 8.2. Section 8.3 presents the summary of the main findings according to the research questions. Section 8.4 outlines the reflections on the conceptual framework. Section 8.5 reflects on the research process, while Section 8.6 presents the conclusions emanating from the study and Section 8.7 outlines the recommendations.

#### 8.2 SUMMARY OF THE STUDY

Educational policies developed in the past (Government of Botswana, 1977) emphasised enrolment at the expense of quality education. The resulting scenario was large class sizes with few resources, culminating in more untrained teachers being employed, resulting in teaching being conducted under unfavourable conditions, with little or no learning materials. During the era of Universal Basic Education, Botswana made significant strides in achieving high school enrolment, with 98% of all primary-school age pupils attending school (MFDP, 2003).

Now the country's emphasis has shifted to quality education with the provision of equity and emphasis on the classroom processes (MoE&SD, 1994). According to Grisay and Mählck (2003), quality in education starts with the development of the relevant learner-centred curriculum, improvement of teacher preparation, improving the methods of teaching and assessing pupils, and provision of resources (Kellagan & Greaney, 2003). When resources are in limited supply, performance assessment suffers because all available resources are channelled towards cost-effective paper-and-pencil tests and measure of low-level content, with the more thought-provoking, complex and abstract content not being evaluated.



It is against this background that the aim of the study was to understand and explore the characteristics and quality processes needed in the performance assessment of agriculture Form Four students, to ensure valid and reliable examinations in Botswana.

The overall research questions which guided the study were:

1. *How valid and reliable are the performance assessment processes in Botswana schools?*
2. *How can quality assurance processes be developed in order to produce valid and reliable marks for BGCSE Agriculture performance assessment?*

To understand effectively the current performance assessment processes in Botswana schools, and to develop quality assessment processes, the main research questions were broken down into sub-research questions. The first main research question was achieved through three approaches, namely; literature review on the current policy on performance assessment; literature review on the conduct of performance assessment internationally; and a baseline survey to understand teachers' practices. The second main research question was addressed by a developmental research approach involving distinct stages of design, formative evaluation, and revision of successive prototypes. Four prototypes were produced and throughout the design and development, practitioners and experts involvement was of paramount importance.

### 8.3 SUMMARY OF THE MAIN FINDINGS

A summary of the main findings is presented, based on the research questions.

#### ***8.3.1 How is performance assessment currently conducted in Botswana schools?***

A survey conducted in two regions purposively sampled, to understand the processes of performance assessment in schools, revealed that teachers were well trained in pedagogical approaches but lacked training in assessment strategies. Teachers' workload was made higher by the large class sizes of up to 50 students. The product assessment was found to be inappropriately carried out as there were no standard criteria used throughout the country to ensure fairness. Each school devised its own assessment criteria based on the syllabus statements and how it interpreted them. Due to lack of standardised criteria, assessment

therefore differed from one school to the other. In some cases, paper-and-pencil tests were used instead of observations of students (Airasian & Russell 2008). Records were not properly kept and did not conform to the ISO standards of labelling, retrievability and retention (ISO 2000).

Large number of students coupled with little understanding of how performance assessment is done, resulted in teachers inflating students' marks with the intention to pass them (Grima & Ventura, 2000). Inflation of marks was also promoted by insufficient monitoring and supervision. As a result, marks collected from schools lacked authenticity. Assessment was largely teacher-centred, with little opportunity given to students to create their own understanding. This de-motivated students to voluntarily choose agriculture in their curriculum, resulting in the majority of student being forced into doing it, hence negative attitude. Involving students in their own assessment allows them to know in advance what and how they would be assessed (Black & William, 1998), and such assessment is important to improve students' learning (Harlen, 2006).

### ***8.3.2 How does the current practice in schools compare with the policy and procedures for performance assessment?***

According to subject groupings by CD&E (see subsection 2.5.3), Agriculture in senior secondary schools was classified under the group of subjects known as *Creative, Technical and Vocational subjects* (MoE&SD, 2002b). However, as far as pedagogy was concerned, agriculture was considered as a 'Full Classes' subject. Being a 'Full Class' subject meant having a minimum number of learners of 30, as per the Revised National Policy on Education of 1994. It was the only subject in the *Creative, Technical and Vocational subjects* grouping, which had the minimum number of 30 learners, while other subjects in the same grouping had a maximum number of 20. Consequently, it had the largest number of students among the optional subjects. The RNPE recommended the maximum<sup>9</sup> number of students in a class to be 35, but was silent on the maximum possible number. Large class sizes make assessment difficult (Jones, 2006) since teachers are forced to assess many students in a limited time. Consequently, majority of teachers resort to assessing products leaving other important aspects of the students' ability that have momentary evidence not assessed (Black, 1995; William & Black, 1996).

---

<sup>9</sup> Agriculture by being classified as a 'full Class' subject meant it could have as many learners as any so called academic subjects such as History, Geography, and English.

Implementing performance assessment in Botswana schools is a challenge as teachers are inadequately trained to conduct performance assessment. Each school, and sometimes each teacher, develops own performance tasks, because the provided criteria is ambiguous and interpreted differently. Ministry officials rarely visited school to monitor implementation and assist teachers.

### ***8.3.3 How does Botswana's experience compares with the international practice?***

The conduct of performance assessment in Agriculture in Botswana is faced with numerous problems as a result of quality assurance processes not being entrenched in the system. Tasks implemented were designed by individual teachers, supervision was insufficient, teachers were not trained to conduct of performance assessment, administrators seemed not to understand their roles clearly, students motivation was low, resources were not enough for all students, and workload was high for teachers.

The moderation of agriculture was done by one visiting moderator at the end of the year, to ratify the teachers' marks on the project report. Anecdotal evidence suggests that there was always friction between teachers and the moderators. This is because the moderators enforced their verdict rather than to reconcile the differences between them and teachers (Radnor & Shaw, 1995). In developed countries, a number of moderation strategies aimed at assuring quality are applied, and inspection is done throughout the year (Boustead, 2008; Lennox, 2000).

The contribution of performance assessment in Agriculture to the final grade was low (20%) due to the difficulties of ascertaining the validity and reliability of performance assessment. It has emerged that this was true as some teachers gave marks without conducting the assessment. The contribution of CA in other countries is high, ranging from 20% to 100% (Broadfoot, Gaseman, 1993; 1994; Raiwoce & Pongi, 2000). The technical adequacy of validity and reliability at international level is assured by embedding quality into the processes (Campbell & Rosznyai, 2002; Richard, 1993), through teacher training to assess; provision of resources; standardising tasks and assessment; accreditation of the school to implement assessment; development of learner support materials; monitoring and supervision; and multiple rating (Chong, 2009; Khoo & Idrus, 2004).

#### ***8.3.4 How can quality assurance processes for performance assessment be developed to ensure valid and reliable marks?***

The development of the quality processes was based on producing the standard tasks and assessment materials to be implemented in a system entrenched with quality assurance processes. The development of the tasks was iteratively done in collaboration with stakeholders. The developed tasks outlined criteria to be met, resources needed, supervision needed and how the marks could be authenticated. The development of the first two prototypes was thus aimed at addressing the validity of the tasks, while the last two prototypes were aimed at achieving the practicality and effectiveness.

The developed tasks were evaluated by experts, teachers and students through questionnaires, interviews and observation and their input was incorporated in the design of the subsequent prototypes. Interviews were done on a one-to-one for clarity except with students. The tasks were found to be well developed and appropriate for the level intended. Teachers applauded the exemplar assessment materials to have been thoughtfully developed; hence the structure and the language were clear to both teachers and students. The quality assurance processes associated with the intervention such as detailed criteria, supervision during implementation of the task, training received on the implementation of the task, and openness of the task were viewed to enhance the validity and reliability of performance assessment (Broadfoot, 1994; Queensland Studies Authority, 1998). All these facilitated assessment of students' skills that was never assessed before such as processes and dispositions.

The intervention also motivated students to work hard to achieve and take responsibility of their learning as they knew in advance the underlying purpose for assessing and what was expected from them (Salvia and Ysseldyke, 1998). They gained confidence in dealing with problem-based learning, collaboration and cooperation in learning ill-structured content (Burris & Garton, 2007) resulting in improved validity of performance assessment.

#### ***8.3.5 What are the characteristics of an effective quality assurance system for ensuring valid and reliable performance assessment nationally?***

The development of the standard tasks and assessment materials was regarded as an important strategy for improving the reliability and validity of assessment process. As indicated before, the prototypes of the materials were development iteratively in collaboration with practitioners, with successive formative evaluation at the end of each cycle. Feedback

from practitioners was incorporated into the redesign and development to ultimately come up with a product that would improve both the assessment practice and learning outcomes. Though the final prototype could not be field tested in real situation for efficiency, the following characteristics can be outlined about the practicality of the standard tasks:

*Assessment of product, processes, and dispositions:* The assessment tool compels teachers to assess processes, products and dispositions (Black, 1995; William & Black, 1996). Teachers have to assess students as they are working. Students can also assess themselves and compare their assessment with that of the teacher. Assessment of processes prevents the possibility of buying products and presenting them for assessment. It also offers the opportunity for those students who are gifted at manipulation to be credited (Ryan & Miyasaka, 1995).

*Openness of task:* tasks were open to offer choice and cater for different developmental levels. Problems in agriculture cannot be conducted under standardized conditions or they do not manifest themselves always in the same way at anytime across context. In addition open tasks allow students to work at their own rate.

*Developing tasks of equivalent demands:* Tasks of equivalent demands (Keightley & Coleman, 2002) ensure that all students are assessed on almost the same skills and activities, using clearly defined criteria. Teachers understanding and interpretation is common thus improving the validity and reliability of assessment.

*Provision of resources:* the implemented tasks revealed that when resources are availed, they helped accelerate the rate at which tasks were assessed, making large classes more manageable and outcome more dependable (Jones, 2002).

*Training of teachers:* the implementation of the standardised tasks was preceded by training of teachers. This helped them to understand the criteria the same and reliability of scoring students improved (Maxwell, 2004).

*Multiplicity of assessments:* the use of more assessors and assessment of the student more than once on the same skill improved reliability of scoring (Thorndike & Thorndike-Christ 2010). This was coupled with feedback which helped students improve on their weakness.

#### 8.4 REFLECTIONS ON THE CONCEPTUAL FRAMEWORK

The conceptual framework for this study is presented in Figure 3.1 (below). It draws heavily on cognitive and constructivist learning theories which purport that learning “requires the active engagement of learners and is determined by what goes on in their minds (James, 2006, p. 55)”. The framework consists of system-level and school-level factors for improving the validity and reliability of performance assessment in Agriculture for certification (See Section 3.7). The former are identified as performance assessment policy; provision of resources; teacher competency; monitoring and supervision; use of standardised materials; teacher workload; and teacher/student ratio. School-level factors are leadership; learning autonomy; student motivation; multiple modes of assessment; multiple rating and student readiness; school leadership; learning autonomy; monitoring and supervision; student motivation; multiple modes of assessment; multiple rating; and student readiness. School administration has the control over school level factors and it is its responsibility to ensure that these factors do not impede the performance assessment process capability (Mamary, 2007; Wiggins, 1998).

The conceptual framework takes cognisance of the fact that learning is a social construction (Broadfoot & Torrance, 1999) which takes into account prior knowledge to be an important determinant of a student capacity to learn new materials (James, 2006). Whenever curriculum is based on these theories learning affords students the opportunity to interact with the other stakeholders among them teachers, peers, parents, social workers, school administration, and school counselors (Salvia & Ysseldyke, 1998) and learn from them. Formative assessment is an important integral component of the pedagogical practice (James, 2006) and has been found to play a significant role in authenticating assessment (Torrance & Pryor, 1998). Resent research summarized by Black and Wiliam (1998) shows that student self-assessment skills, learned and applied as part of formative assessment, enhances student achievement (Torrance & Pryor, 1998). Assessment is thus an ongoing process aimed at understanding and improving student learning (Angelo, 1995).

Leaning underpinned by cognitive and constructivist theories facilitates inclusion of more demanding tasks of investigation, problem solving, report-writing, etc in the curriculum, and in turn have to think of more flexible ways of assessing such activities than traditional paper-and-pencil tests (James, 2006). As Resnick and Resnick (1992) have put it: ‘if we put

debates, essays, discussions and problem solving into testing system, students will spend time practising those activities” p 59. Students could learn more if they are assessed appropriately using the appropriate methods and by capable assessors.

For the system to produce valid and reliable marks, quality assurance processes have to be embedded, both at school level and system level. Among the system-level and school-level factors, findings revealed that there were those which were more important than others for the successful conduct of valid and reliable performance assessment marks at both system and school levels (See Figure 8.1, above). These were labelled as ‘principal factors’. If they were not present or provided for, the probability of producing valid and reliable marks was minimal. Under system level, such factors are monitoring and supervision; the availability of assessment policy; provision of resources; use of standardised materials; and teacher competency. The minor system-level factors were teacher/student ration and teacher workload. Once the principal school-level factors were satisfied, minor ones followed, as they were to a large extent dependent upon them. On the other hand, principal school-level factors are multiple modes of assessment, monitoring and supervision, and multiple rating, while minor school-level factors are learning autonomy, student motivation, and student readiness.

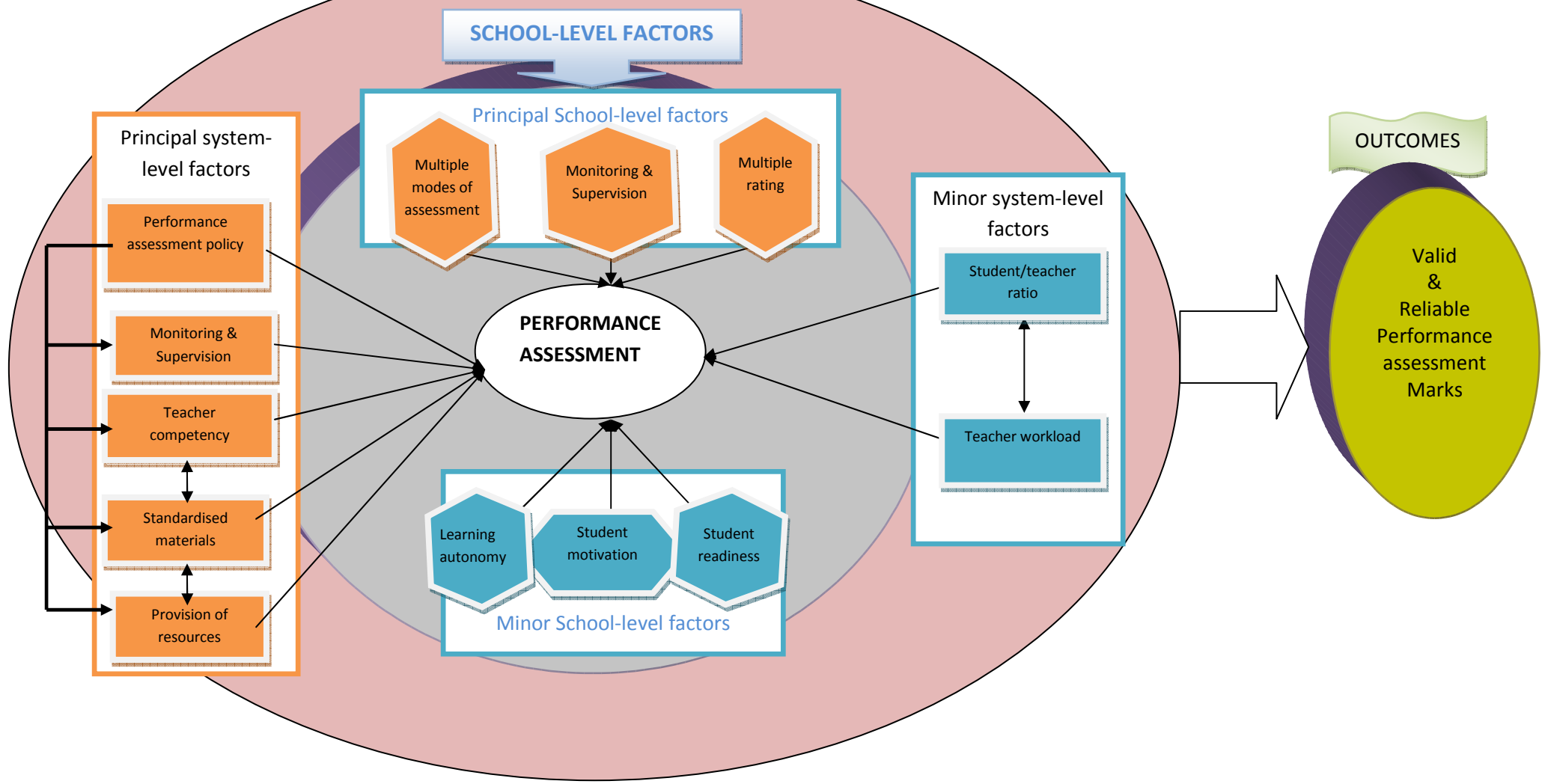


Figure 8.1: Characteristics and quality processes affecting validity and reliability of performance assessment marks



In most cases, system-level factors are outside the control of schools, and determined by the Ministry of Education officials. However, that does not mean that schools and teachers can remain helpless when the situation deteriorates, waiting for the Ministry officials to act. For example, teachers can upgrade their level of performance assessment on their own, or schools can raise funds to buy tools and equipment. The use of standardised materials proved to be extremely useful in obtaining valid and reliable performance assessment marks. With both the teachers' understanding the objective of assessment, using standard criteria, and their knowing each party's expectation, they were motivated to benefit most from assessment (Black & William, 1998). Students in particular wished to use criteria to evaluate their own work prior to the teacher's evaluation.

The development of standardised materials was closely related to teacher competency and provision of resources. For teachers to produce quality materials, it was found that they needed to be grounded in assessment methodologies (Popham, 2005). Development of assessment tasks should be done in consultation with students (James, 2006; Mergendoller, Markham, Ravitz, & Larmer, 2006; Stiggins, 1997), who ultimately should have the right to the assessment procedures and be willing to complete them (Wiggins, 1998). Assessment should not be forced upon students as this might lead to wrong assessment data being collected and used to make improper decisions.

Student/teacher and teacher workload were also related, as the higher the student/teacher ratio the higher the teacher workload. However, as indicated above, if all principal factors were provided for, the minor factors' effects were not felt. The system and school efforts should be directed towards availing the principal factors.

Monitoring and supervision constituted a subset of both system-level and school-level, and was placed between the two. Senior teachers and school administrators, when they monitored and supervised, provided evidence of improvement in the outcomes (Mamary, 2007). Ministry officials also conducted spot-check visits to ascertain that performance assessment had been conducted properly.

Although policy formulation is grouped together with other principal factors, it weighed the most, because everything emanated from it. Policy is needed to guide schools on how to conduct performance assessment, who should conduct performance assessment, how many tasks should be done, what is the role of the student, what resources schools should have,

how the marks should be stored, and what supervision is needed. Student/teacher ratio and impact of teacher workload became negligible when the primary factors were in place, because time which was identified as an important resource no longer became an issue when everybody had tools to work with. Students had access to the assessment guide and were aware of what was expected, as well as understanding and valuing the importance of performance assessment.

Resources provision is an important factor for the implementation of performance assessment (Maxwell, 2004), and resources in Agriculture were identified as time, multiple assessors, tools and equipment, standardised tasks and assessment materials. Since the developed standardised materials enumerated all prerequisites needed for the successful implementation of the practical, this helped in reducing the time needed for assessment as materials and tools were organised in advance, saving time for assessment purposes, even for large class sizes. Thus, constructivist strategies to learning normally associated with small class sizes, are such as cooperation, problem-based learning, discussion, discovery, scaffolding, and collaboration (Gronlund, 2003; James, 2006).

Whenever these strategies were applied, they facilitated students' engagement in active construction of own knowledge (Eysenck, 2004; Slavin, 1994) and displayed growth in problem solving skills (Burriss & Garton, 2007). As noted by Johnson et al. (2009), Mills (1996), and Nitko and Russell (2007), teaching large classes with limited resources impacts on the time available, resulting in failure to offer individualised instruction. For example, studies have found that students exposed to problem-based learning consistently performed better.

The use of more assessors and multiple assessments resulted in a more acceptable mark by the students, as they believed that the second assessor neutralised bias. Students' acquisition of knowledge and skills cannot be adequately and comprehensively measured by a single mode of assessment as there are different kinds of achievement to assess (Stiggins, 1997). Airasian (2005), Mamary (2007) and Maxwell (2004) assert that assessment that is fair, leading to valid inferences with minimum error, is a result of a series of measures using various assessment methods that show student understanding through multiple methods in a variety of contexts or settings, rather than just administering a test.

## 8.5 REFLECTIONS ON THE RESEARCH APPROACH

### *8.5.1 Methodological reflections*

This study employed the design research or development research approach as outlined in Chapters 1, 4 and 6. Design research was appropriate because it made possible the identification of the root cause of the problem in performance assessment, description of performance assessment practices and processes, and obtaining points of views and attitudes held by practitioners (Barab & Squire, 2004; Kelly, 2004; Persse, 2006), through a baseline survey. Based on the findings of baseline survey, design research allowed for designing and developing the intervention in collaboration with practitioners and other stakeholders in education. These were involved at various stages of the design and development process (Barab & Squire, 2004; Kelly, 2004), adopting a cyclic approach of design, evaluation and revision (Plomp 2008; Van den Akker, Branch, Gustafson, Nieveen & Plomp, 1999).

Data collection and analysis employed a mixed method approach whereby both quantitative and qualitative methods were used. During the baseline survey, a variety of data collection instruments developed by the researcher were employed, such as teacher questionnaire, administrator questionnaire, teacher interview schedules, and document analysis. Thus, triangulating of data sources helped in improving the validity and reliability of information collected (Mertens, 2010) and resulted in rigorous, empirically grounded claims and assertions (Cobb et al., 2003). These instruments were reviewed by experts before piloting. The questionnaires were handed to the sampled schools by the researcher and interviews arranged for later dates. Purposive sampling was used to get an in-depth understanding of the phenomenon by identifying information-rich participants (Mertens, 2010).

The outcomes of the baseline survey guided the design and development of the exemplary assessment materials that were implemented by teachers during the intervention phase. The intervention phase employed multiple design-test-revise cycles in the interactive and iterative development of standardised performance tasks and assessment materials aimed at improving the quality of the outcomes (Barab & Squire, 2004; Collins et al., 2004). The review of the first prototype was achieved through administration to experts of a questionnaire designed specifically for checking content, design, and technical quality, according to Tessmer's layers of formative evaluation presented in Figure 4.4.

During the implementation of the second and third prototypes, the researcher inducted teachers to the developed materials in a workshop so as to give them the same understanding. They in turn explained to their students what was expected of them. The evaluation comprised observation of teachers and their students during the conduct of performance assessment for practicality. Both teacher and student questionnaires and interviews were also administered. Teachers were highly resistant to implementing the intervention, yet during the interviews they indicated that they liked the intervention. One teacher during the piloting of the intervention had to be followed by the senior teacher to at least assist the researcher.

The study remained flexible throughout, to accommodate the ever-changing nature of natural settings (Mertens, 2010; Ornstein & Hunkins, 1993). It must be mentioned that due to triangulation of data collection, a lot of data was produced, which presented a challenge to the researcher on how to handle it. Nevertheless, the combination of qualitative and quantitative methods of data collection and analysis led to better understanding of the characteristics of the quality assurance processes needed for implementing performance assessment in agriculture at Form Four level (Barab & Squire, 2004; Collins et al., 2004). Another problem associated with this design research was lack of generalisability to similar situations, due to a purposive sampling technique employed. The nature of design research is such that it should culminate in the intervention being ultimately field-tested in real world settings to test for efficiency. Due to teachers and their Trade Unions engaging in industrial action during the examination time, the final and fourth prototype was not summatively evaluated, hence the conclusion about the intervention's effectiveness could not be made.

### ***8.5.2 Reflection on researcher's role***

The nature of the design research meant that the researcher had to assume multiple roles of designer-developer, facilitator and evaluator during the study. The researcher designed and developed the initial exemplar assessment materials in collaboration with practitioners and experts. He went on to facilitate teachers on how to implement the developed exemplar assessment materials, then observe them implementing the intervention. During the design process, the researcher was concerned with producing high quality materials. As a facilitator, the researcher aimed at ensuring that teachers understood the way the intervention was to be implemented, while as an evaluator, the researcher was to be as objective as possible.

Playing multiple roles in the same study was both beneficial and problematic. It was beneficial in the sense that the researcher interacted with practitioners who understood the root cause of the performance assessment problem, aiding in designing an intervention that was appropriate to solve the problem of performance assessment, and design principles that characterise the intervention (Cobb et al., 2003; Collins, Joseph & Bielaczyc, 2004; Gravemeijer & Cobb, 2006; Plomp & Nieveen, 2007).

As indicated in Subsection 4.6.2, the researcher was well known to participants, since he had worked with them as a teacher and later as an Education Officer. This could have had an effect on the way teachers perceived the researcher, as an Education Officer who was inspecting their individual assessment practices rather than as a researcher auditing the assessment process. As a result, the outcome could have been biased. To ensure that inferences made from the information collected were valid and reliable, and to check for consistency of evidence, triangulation of data collection instruments and sources was employed (Mertens, 2010). The researcher, as the main qualitative data collection instrument, was sensitive, adaptable and responsive to changing circumstances, and posed as a non-participant observer (Patton, 1990) so as not to influence the outcome.

## 8.6 CONCLUSIONS

The following are conclusions drawn from this study:

**The validity and reliability of performance assessment in agriculture in Botswana schools needs to be improved.**

The validity and reliability of performance assessment of agriculture in Botswana schools needs to be improved through a number of factors. Firstly a policy on performance assessment needs to be formulated to guide the conduct of performance assessment for certification. Because of the absence of a policy processes and practices varied from school to school and from one teacher to another, resulting in varying standards. Since there were no standard tasks developed, teachers developed their own, of different demands and quantity. The official document provided by the Ministry of Education, as pointed out in Chapter 2, was not clear or detailed and provided room for such variations. It was found that performance assessment was conducted by teachers with little or no training, and only a few

were inducted on the conduct of performance assessment when they joined the profession (Subsection 5.2.3). Lack of training resulted in below standard teachers' performance assessment practices (Section 5.4), as evidenced by teachers' assessment based mainly on products and little on processes or affect. This disadvantaged those students who were not good at producing a product (Ryan & Miyasaka, 1995) and credited those who presented a product irrespective of how it was produced.

Furthermore, successful implementation of performance assessment depends on the availability of resources (Maxwell, 2004). It is outlined in Chapters 5, 6 and 7 that resources to conduct performance assessment in schools were insufficient. The most important resource found to be in acute shortage for individualised assessment was time, which was a function of other resources (human, physical and workload). Insufficiency of other resources impacted on the time to conduct appropriate performance assessment. For example, teachers using their own developed assessment materials in poorly resourced large classes, could not assess all students, compelling them to devise other means to assess, which led to invalid and unreliable outcomes, as outlined in Chapter 5.

Monitoring of performance assessment was insufficient, giving some teachers the opportunity to alter students' marks.. This resulted in performance assessment marks being unauthentic, thus requiring thorough supervision and monitoring at both school-level and system-level.

**The absence of policy and clear guidelines for performance assessment results in variable and sometimes inappropriate implementation at school level.**

Teachers were found to employ only one mode of assessment, namely product assessment, at the expense of processes, and affect (Section 5.4). The absence of standard criteria to be used throughout the country resulted in product assessment being inappropriately carried out, despite it being more objective than process assessment. Each school devised its own assessment criteria based on its interpretation of the available unclear criteria provided in the syllabus (see Section 2.8). Because of the absence of clear policy and criteria, any teacher could assess, irrespective of training background in performance assessment. Teachers therefore assigned inflated group scores with the aim of passing students since marks were used for certification (Sections 1.3 and 2.8).

Assessment of students was teacher-centred, with little emphasis placed on student autonomy in learning (Subsection 5.3.2), and secretly conducted. Students were not informed in

advance and even if they were, it was not clear how it would be conducted. This discouraged students who developed negative attitude towards the subject (Sections 5.3 and 5.4). Studies show that involving students in their own assessment allows them to know in advance what and how they would be assessed (Black & William, 1998), and they can use the criteria to evaluate their own work prior to the teacher's evaluation, in turn leading to improvement in learning (Harlen, 2006).

**Performance assessment practices for agriculture in Botswana schools are not up to standard when compared to international best practices.**

The conduct of performance assessment in Botswana schools, as pointed out in Subsection 5.3 is even done by teachers who have not trained in performance assessment. Training of teachers to acquire the appropriate expertise is essential (Broadfoot, 1994) and is emphasised in developed countries (Maxwell, 2004; Queensland Studies Authority, 1998). Assessment procedures are largely the responsibility of teachers, even for certification and selection purposes, with minimal external intervention or moderation (Gasemann, 1993).

International best practice requires multiple raters and multiple rating of a student work. But in Botswana, only one rater is used and students are not given the second chance when they did not achieve. In developed countries inter-rater reliability of the moderation system for performance assessment surpasses that of many external examination regimes. The kind of moderation applied is the one directed at ensuring quality (Boustead, 2008; Broadfoot, 1994; Harlen, 1994; Maxwell, 2004; Raivoce & Pongi, 2000). However, the use of a variety of methods that combine both quality assurance and quality control procedures are also employed and yield better results (Berry, 2008; Keightley & Coleman, 2002; Queensland Studies Authority, 2009; Maxwell, 2004; Raffan, 2000; Raivoce & Pongi, 2000).

School approval or accreditation is seen as an important factor in ensuring quality in performance assessment (Council for Higher Education Accreditation {CHEA}, 2002). Before the school is allowed to conduct performance assessment, a holistic audit of its capabilities to successfully implement performance assessment is conducted (CHEA, 2002; Colbeck, Caffrey, Donald, Lattuca, Reason, Strauss, Terenzini, Volkweinm, and Reindl, 2000; Jones, 2002). Schools are required to submit a detailed assessment programme and procedures (Grima & Ventura, 2000; Keightley & Coleman, 2002; Raivoce & Pongi, 2000). Regular visits are made throughout the conduct of performance assessment to verify that

internal assessment programmes are being followed and to assist teachers in the delivery of the learning programmes (Keightley & Coleman, 2002). None of these are done on Botswana. Schools are not accredited, no detailed plan of assessment is required and school visits are infrequent. All is left to the teacher, senior teacher and school administration whom it was found none of the parties executed its responsibility well.

**The use of agriculture standard tasks and assessment materials developed collaboratively with students, practitioners and experts at both school level and system level may lead to improvement in quality assurance and student outcomes**

The involvement of practitioners, students and experts in the development of standard tasks and assessment materials is important to situate the learning requirement to the level of the learner. According to the theory of constructivism, learning is determined by what goes on in people's mind (James, 2006. p.55). Involving students in the development of materials for learning and assessment provides them the opportunity to identify some shortcomings in the materials developed. Stakeholders were able to identify, Issues identified were such as physically inaccessibility and activities targeting students' of different abilities which were included in the subsequent prototypes. The stakeholders brought different expertise and experiences which helped refine the standard tasks and assessment materials. Stakeholders identified gaps in the developed materials, such as failure to provide opportunities for all students to interact and cooperate within a group, lacking assessment of affective domain, lack of objectivity in scoring, and language being unclear (Section 6.4).

Teachers' instructional practices and knowledge of assessment improves significantly when they were involved in developing successive prototypes of the interventions. For example, teachers initially concentrated on assessing products, thus showing deficiency in skills to assess other aspects of performance. They initially resisted change because of insufficient knowledge of the proposed change (Harrington, McElroy & Morrow, 1990) and the value change was likely to bring (Kent & McNergney, 1999), as well as ingrained personal teacher's beliefs (Ertmer, 1999). Eventually, teachers were comfortable in handling processes and affect assessment through the use of detailed easy-to-use assessment criteria.

The involvement of students in assessment adds value, as viewed by both teachers and students, and consequently makes the teachers' work easier. Students' record-keeping and its importance were enhanced, and teachers' scoring using the criteria (6.11.4) resulted in



objective scoring; enhancement of maintenance of standards; increased motivation of teachers and students and a change in students' perception of performance assessment from negative to positive, as they proactively took responsibility for their learning (Black & Wiliam, 1998; Harlen, 2006; Salvia and Ysseldyke, (1998).

**The main characteristics of an effective system for ensuring valid and reliable performance assessment in Botswana for Agriculture comprise of system-level and school-level factors (Redo)**

An effective system for ensuring valid and reliable performance assessment for Agriculture in Botswana schools should include clearly written assessment policy, well-trained teachers to assess, sufficient resources, close monitoring and supervision, and the use of standard tasks and assessment materials. The policy is the foundation that guides all the activities associated with performance assessment. These include how assessment should be carried out, who should assess, how many tasks should be assessed, the weight of assessment, resources needed for performance assessment, inspection of schools, and students responsibility. Assessment policy can also be used as a tool for defence during litigation.

Trained teachers in performance assessment apply student-centred approaches to assessment such as formative assessment or assessment for learning which draw on cognitive and constructive theories of learning (James, 2006). These results in improved learning (Black & Wiliam, (1998a, 1998b; Izard, 1998), because students are consulted in decision-making concerning students' assessment, approach assessment as an open transaction aimed at improving students' learning instead of auditing their knowledge, and apply multi-modal and multiple assessments.

Performance assessment should be carried out in well resourced schools, which should be approved to conduct performance assessment after thorough inspection of their resources. Well-resourced schools have been found to perform better (Howie & Plomp, 2001) because teachers can facilitate collaborative working, individualised assessment, multiple assessment and reassessment. Although workload had been found to be a hindrance to effective conduct of performance assessment (Howie, 2006; Howie & Plomp, 2003), its effects are diluted when principle factors are availed.

Both internal and external monitoring and supervision on the conduct of performance assessment are important to ensure that teachers adhere to standards. To ensure that

monitoring was effected, standardised tasks were developed with the provision for teachers and senior teachers to sign as a way of certifying that assessment had been conducted according to the appropriate standard. This compelled assessment of processes to be done while students were working. Internal monitoring should be frequent and ultimate responsibility lies with the school head as the overseer of the school activities (Mamary, 2007; Wiggins, 1998). Singh, (2000) posits that monitoring by external officers from the Ministry should be strategic and random, not only to find faults but also to support teachers in implementation.

The use of standardised tasks and assessment materials by teachers in turn standardises assessment and provides valuable information for further intervention (Mamary, 2007). In-service training should be organised to impart teachers with skills to develop sound assessment instruments (Chong, 2009; Halsall, 1998; Maxwell 2004; McMillan, 2004; Popham, 2005).

## 8.7 RECOMMENDATIONS

The conclusions based on the findings for this study have highlighted some important issues that need to be followed up in order to improve the performance assessment of Agriculture for certification in Botswana schools. The first step in ensuring quality in performance assessment is to improve quality of the processes (Campbell & Rosznyai, 2002; Richard, 1993). Improvement should be directed towards teachers' assessment skills; resources provision to schools; the development and use of standardised tasks and assessment materials, strengthening monitoring and supervision; multi-rating, the use of multi-modes of assessment and accrediting schools to conduct performance assessment (Chong, 2009; Khoo & Idrus, 2004). Recommendations emanating from this study are therefore grouped into policy, training and development, practice, and further research.

### 8.7.1 Policy

The Ministry of Education and Skills Development has long recommended the introduction of continuous assessment (CA), of which performance assessment is a component (RNPE, 1994). However, this aspect of the policy has not yet been fully implemented, after seventeen years. This is because teachers in different schools and even in the same school still do not

conduct standardised performance assessment tasks due to lack of policy and guidelines. For effective implementation of performance assessment for certification, there is a need for a written policy to guide practice. The policy should clearly spell out:

- The roles and responsibilities of different departments of the Ministry of Education and Skills Development who are important stakeholders in performance assessment as discussed in Section 3.5.
- The conditions under which performance assessment should be conducted (James, 2006) and the roles of players within the school set up. There is confusion and lack of clarity on roles among professionals (Chong, 2009), since it is not documented as to which teachers should conduct performance assessment for certification and which should not. This resulted in everybody conducting it, even those not trained. The consequence of this was lowering of the weight of performance assessment due to claims of low validity and reliability.
- The number of tasks that the student should be assessed in each content domain and how that should be done.
- Tasks and assessment materials should be standardised, since there were neither exemplar assessment materials developed nor guidelines on how to develop them. Teachers developed their own tasks, which significantly differed in demands resulting in unfair assessment.

### ***8.7.2 Training and development***

Teachers conducting performance assessment are not trained to assess (Subsection 5.2.3). Teachers should be given adequate training so that they effectively implement performance assessment for certification. Once teachers are trained, they can develop their own performance assessment tasks to assess what is inaccessible to external examination (Pellegrino, Chudowsky & Glaser, 2001; Tindal & Haladyna, 2002). Since learning is socially constructed (James 2006), training institutions should design and develop a course in collaboration with the stakeholders to be offered to pre-service student-teachers. Similarly, a relevant course tailor-made for in-service teachers should be developed. As once noted by Nitko (1998), officers from the Department of the Ministry of Education charged with the responsibility of monitoring and supervising performance assessment should be also trained on the conduct of performance assessment.

### 8.7.3 *Practice*

Standardised tasks and performance assessment materials should be developed for use by teachers. They should be developed by the responsible Department of the Ministry of Education and Skills Development in collaboration with practitioners and experts. The tasks should, inter alia, be student-centred to allow students to create knowledge; to be open to cater for differential learning rates; allow for multiple assessments; encompass all domains of ability; incorporate high order thinking skills; and be physically assessable to all students. Development must proceed by identifying those objectives from the syllabus which lend themselves to performance assessment from each topic, followed by developing equivalent tasks. This should be run parallel with training teachers on the implementation of performance assessment tasks.

Monitoring and supervision of performance assessment is not sufficient. This was probably due to lack of role clarity as well as training (Subsection 5.2.3) among senior teachers, school administration and the Ministry of Education and Skills Development. This raised doubt as to whether this kind of assessment was given the same consideration as the paper-and-pencil testing. To strengthen the monitoring and supervision at school level, the Ministry of Education and Skills Development should consider the establishment of a fully fledged Quality Assessment and Assurance Department (QAAD), headed by a qualified teacher in assessment issues at the post of head of department (HOD). Because quality assurance can only work with total senior administration commitment, administration should be thoroughly inducted on the formulation, implementation and review of a quality policy (Richard, 1993).

Provision of resources was found to be of paramount importance in the successful implementation of performance assessment. Schools conduct performance assessment with limited resources resulting in outcomes of low validity and reliability. Schools should be accredited to offer performance assessment, which would entail, among others, auditing of the required resources. Time is another important resource fundamental for the success of performance assessment implementation. It should be provided for in terms of reducing class sizes for Agriculture as is the case with other practical subjects.

Agriculture classes were found to exceed the maximum number stipulated in the policy (Section 1.3), thus conferring more work to teachers and limiting the contact time between individual student and the teacher. Agriculture should be reclassified as a “Non Full Class”

(Section 2.6), as with other practical subjects. This would lead to a reduction in class sizes, resulting in manageable student numbers and facilitating the conduct of performance assessment. Reducing class sizes is a practicable possibility as there are many qualified unemployed agriculture teachers (Bennel & Molwane, 2008). Both schools and teachers have to be accredited to offer performance assessment, and the accreditation should be renewed biannually to maintain standards.

#### **8.7.4 Further research**

The study provides evidence that the implementation of quality standardised tasks and assessment materials is one aspect needed for quality assurance processes for improving the validity and reliability of performance assessment outcomes. Students and teachers alike embraced the intervention. However, the study employed a design research which had a limited sample, hence the results could not be generalised to all the schools but provides a fundamental basis upon which further studies could be built. It is suggested that further research be conducted with a larger sample size, which will enable the results to be generalised to all schools and different contexts with confidence.

Although the intervention was welcomed by both teachers and students, the use of summary marksheets is still presenting some problems to teachers during implementation. A summative evaluation is needed to further understand how teachers finally implement it to yield valid and reliable outcomes without requiring more work from teachers. The intervention produces a lot of paperwork, which is of concern to teachers. However, the need for producing records meeting the requirements of labelling, retrievability and retention as demanded by ISO is indisputable (Richards, 1993). Further examination on how this could be achieved without negatively impacting on teachers' morale is imperative.

Investigating all these issues can have an effect of improving agriculture performance assessment to yield valid and reliable marks for certification, as it has been established that performance assessment practices in Botswana schools was inappropriate. However, developing standardised tasks and subsequent training teachers on how to implement and supervisors it, in a well resourced environment, produced valid and reliable outcomes.

## REFERENCES

- Abraham, O. T. (2008). *Oral and Testing in NECO SSCE: Prospects and Challenges*. A paper presented at the 26<sup>th</sup> AEAA Annual Conference. Accra. Ghana.
- Abramowich, E. (2005). *Six Sigma for Growth: Driving Profitable Top-Line Results*. Clementi Loop: John Wiley & Sons Pty Ltd.
- Achilles, C. M. (2005). Class size and learning. In L. W. Hughes (Ed.), *Current Issues in School Leadership*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.
- Agar, M., & Hobbs, J. (1982). Interpreting discourse coherence and the analysis of ethnographic interviews. In Hardy & Bryman (Eds.), *The handbook of data analysis*. London: Sage Publications Ltd.
- Aiken, L. R. (1996). *Rating Scales and Checklists: Evaluating Behaviour, Personality, and Attitudes*. New York: John Wiley & Sons, Inc.
- Airasian, P. W., & Russell, M. K. (2008). *Classroom Assessment– Concepts and Applications*. Boston: McGraw Hill.
- Airasian, P. W. (2005). *Classroom Assessment – Concepts and Applications*. Boston: McGraw Hill.
- Airasian, P. W., & Abrams, L. M. (2002). What role will assessment play in school in the future? In Lissitz & Schafer (Eds.), *Both means and ends*. Boston: Allyn and Bacon:
- Alder, P., & Alder, P. (1994). Observational techniques. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research*. California: SAGE Publications, Inc...
- Alreck, P. L., & Settle, R. B. (1995). *The Survey Research Handbook: Guidelines and Strategies for Conducting a Survey*. New York: McGraw-Hill.
- Altheide, D., & Johnson, J. (1994). Criteria for assessing interpretive validity in qualitative research. In N. Denzin & Y. Lincoln (Eds.), *Handbook of Qualitative Research*. London: Sage.

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Angrist, J. D., & Lavy, V. (1999). Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement. *Quarterly Journal of Economics* 114(2): 533-575.
- Ary, D., Jacobs, L. C., Razavieh, A., & Sorensen, C. (2006). *Introduction to Research in Education*. Australia: Thompson Wadsworth.
- Arter, J., & McTighe, J. (2001). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance*. Thousand Oaks, CA: Corwin Press.
- Assessment Reform Group (2002). *Testing, Motivation and Learning*. ARG, Cambridge.
- Assessment Reform Group (2006). *The Role of Teachers in the Assessment of Learning*, ARG, Cambridge.
- Babbie, E., Halley, F., & Zaino, J. (2000). *Adventures in Social Research: Data Analysis Using SPSS for Windows 95/98*. Illinois: SAGE Publications Ltd.
- Baku, J. J. K. (2008). Assessment for learning and assessment of learning: a search for an appropriate balance. *Journal of Educational Assessment in Africa*. 2(1), 44-52.
- Barab, S. and Squire, K . (2004). Design-based research: Putting a stake in the ground. *Journal of the learning sciences*, 13(1), 1-14.
- Barbour R. S., & Kitzinger, J. (1999). Introduction: the challenge and promise of focus groups. In Barbour R. S., & Kitzinger, J. (eds). *Developing Focus Group Research: Politics, Theory and Practice*. London: Sage Publications.
- Basu, R., & Wright, N. (2003). *Quality Beyond Six Sigma*. Oxford: Elsevier Ltd.
- Bennel, P., & Molwane, A. B. (2008). *Teacher Supply and demand for Botswana primary and secondary schools: 2006-2016*. Gaborone: Government Printer.

- Bennett, J., & Taylor, C. (2004). *Is Assessment For Learning In A High-Stakes Environment A Reasonable Expectation?*. A paper presented at the third Conference of the Association of Commonwealth Examination and Association Boards. Nadi, Fiji.
- Berry, R. (2008). *School-Based Assessment in Hong Kong: Policies, Issues, and Practice*. A paper presented at the IAEA Annual Conference. Cambridge. UK.
- Black, P. (1995). Continuous assessment: Teachers use assessment to improve learning? *British Journal of Curriculum and Assessment*, 5(2), 7-11.
- Black, P. (1993). Formative and summative assessment by teachers. *Studies in Science Education*, 21:49-97.
- Black, P., & Wiliam, D. (1998a). Assessment and Classroom Learning, *Assessment in Education*, Vol. 5, pp. 7-74.
- Black, P., & William, D. (1998). *Inside the Black Box: Raising standards through classroom assessment*, School of Education, King's College, London.
- Boustead, T. M. (2008). *Moderating subject variation within a New Zealand standards-based assessment system*. A paper presented at the 2008 IAEA conference. Cambridge.
- Bogdan, R. C., & Biklen, S. K. (2003). *Qualitative research for education*. (4<sup>th</sup> ed.). Boston: Allyn & Bacon.
- Bordens, K. S., & Abbott, B. B. (2005). *Research Design and Methods: A process Approach*. McGraw Hill: New York.
- Breakthrough Management Group. (2007). *Lean Six Sigma*. New York: Alpha.
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scale. *Journal of Personality*, 54, 106-148.
- Broadfoot, P., (1994). Approaches to quality assurance and quality control in six Countries. In Harlen, W. (Ed.), *Enhancing Quality in Assessment*. London: Paul Chapman Publishing Ltd.



- Brookhart, S. M. (2002). What will teachers know about Assessment and how will that improve Instruction? In Lissitz & Schafer (Eds.), *Both means and ends*. Allyn and Bacon: Boston.
- Burger, S. E., & Burger, D. L. (1994). Determining the validity of performance-based assessment. *Educational Measurement: issues and practice*. 20 (8) 9-15.
- Burris, S., & Garton, B. L. (2007). Effect of instructional strategy on critical thinking and content knowledge: Using problem-Based learning in the secondary classroom. *Journal of Agricultural education*. 48(1), 106-116.
- Butler-Kisber, L. (2010). *Qualitative Inquiry: thematic, Narrative, and Arts-Informed Perspective*. New Delhi: SAGE Publications.
- Camara, W. (2003). *Scoring essay on the SAT written section* (Research Summary No. 10). New York: The College Board. Retrieved June 28, 2009, from [www.collegeboard.com/research/pdf/031367researchsummary\\_26516.pdf](http://www.collegeboard.com/research/pdf/031367researchsummary_26516.pdf).
- Calvo-Mora, A., Leal, A., & Roldan, J. L. (2006). Using enablers of the EFQM model to manage institutions of higher education. *Quality Assurance in Education*. 14 (2) 99-122.
- Campbell, C., & Rosznyai, C. (2002). *Quality Assurance and the development of a course Programme*. Bucharest: UNESCO.
- Chong, K. K. K. (2009). *Whither school-based coursework assessment in Singapore*. A paper presented at the 35<sup>th</sup> IAEA Conference. Australia.
- Christmann, E. P., & Badgett, J. L. (2003). A Meta-Analytic Comparisons of the Effects of Computer-Assisted Instruction on Elementary Students' Academic Achievement. *Information Technology in childhood Education Annual*, 4(1), 91-104.
- Cizek, G. J. (1991). Innovation or enervation?: Performance assessment in perspective. *Phi Delta Kappan* 72 (9) 695-699.
- Clauser, B. E., Harik, P., & Margolis, M., J. (2006). A Multivariate Generalisability Analysis of Data from a Performance assessment of Physician's Clinical Skills. *Journal of Educational Measurement*, 43(3) 173-191.

- Clauser, B. (2000). Recurrent issues and recent advances in scoring Performance assessments. *Applied Psychological Measurement*, 24(4), 310-324.
- Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9-13.
- Cohen, L., & Manion, L. (1989). *Research methods in Education*. London: Routledge.
- Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in Education*. London: Routledge Falmer.
- Comfrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2<sup>nd</sup> ed.). Hillside, NJ: Erlbaum.
- Colbeck, C. L., Caffrey, H. S., Donald, E., Lattuca, L. R., Reason, R., Strauss, L. C., Terenzini, P. T., Volkwein, J. F., & Reindl, T. (2000). *What Works: Policy Seminar on Students Success, Accreditation and Quality Assurance*. U. S. Department of Education.
- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design Research: Theoretical and Methodological Issues. *Journal of the Learning Sciences*. 13(1), 15-42.
- Colton, D., & Covert, R. W. (2007). *Designing and Constructing Instruments for Social Research and Evaluation*. San Francisco: John Wiley & Sons, Inc.
- Coolican, H. (2006). *Introduction to Research Methods in Psychology*. London: Hodder Arnold.
- Council for Higher Education Accreditation. (2002). *The Fundamentals of Accreditation-What do you need to know?* Washington DC.
- Cresswell, J. W. (2009). *Research design: Quantitative, qualitative, and mixed methods approaches*. (3<sup>rd</sup> ed.). Thousand Oaks, CA: SAGE.
- Crooks T. (2004). *Tensions between assessment for learning and assessment for qualifications*. Paper presented at the Third Conference of the Association of Commonwealth Examinations and Accreditation Bodies (ACEAB) Nadi, Fiji, 8-12.

- Dancy, P. & Reidy, J. (2002). *Statistics Without Maths for Psychology*. London: Pearson Education.
- Deakin Crick, R., Broadfoot, P., and Claxton, G. (2003). *Developing the ELLI: The Effective Lifelong Learning Inventory in Practice*. Graduate School of Education, University of Bristol.
- De Vellis, R. F. (2003). *Scale Development: Theory and Practice*. California: Sage Publications.
- Devitt. H. J., Kurrek. M.M., Cohen. M. M., & Cleave-Hogg, D. (2001). *Anesthesiology*. 95 (1), 36-42. Retrieved 10 June, 2009 from <http://journals.iww.com/anesthesiology/pages/articleviewer.aspx?year=2001&issue=0>
- Dick, W., Carey, L. & Carey, J.O. (2009). *The Systematic Design of Instruction*. New Jersey: Pearson.
- Diez, M. E. (2002). How will teacher education use assessments? An assessment scenario from the future. In Lissitz & Schafer (Eds.), *Assessment in Educational Reform: both means and ends*. Boston: Allyn and Bacon.
- Doty, L. A. (1996). *Statistical Process Control*. (2<sup>nd</sup> ed.). New York: Industrial Press Inc.
- Donald, D., Lazarus, S., & Lolwana, P. (2002). *Educational Psychology in Social Context*. Cape Town: Oxford University Press.
- Doty, L. A. (1996). *Statistical Process Control*. 2<sup>nd</sup> Edition New York. Industrial Press Inc.
- Downing, S. (2006). Twelve steps for effective test development. In S. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 3-25). Mahwah, NJ: Erlbaum.
- Driscoll, M. P. (2000). *Learning for instruction*. Massachusetts: Pearson Education Company
- Dyer, C. (1995). *Beginning Research in Psychology: A Practical Guide to Research Methods and Statistics*. Oxford: Blackwell Publishers.
- Eckes, G. (2003). *Six Sigma for Everyone*. New Jersey: John Wiley & Sons, Inc. Remember, people are rarely the root cause of a problem in a process.

- Ertmer, P. (1999). Addressing first- and second-order barriers to change: Strategies for technology integration. *Educational Technology Research and Development*, 47(4), 47-61.
- Eysenck, M. W. (2004). *Psychology: An International Perspective*. New York: Psychology Press Ltd.
- Field, A. (2000). *Discovering statistics using SPSS for windows – Advanced Technique for the Beginner*. London: SAGE Publications.
- Fink, A. (1995). *How to Analyze Survey Data*. California: SAGE Publications, Inc.
- Fink, A. (2005). *Evaluation fundamentals: Insights into Outcomes, Effectiveness and Quality of Health Programs*. Thousand Oaks: SAGE Publications, Inc.
- Fink, A. (2009). *How to Conduct Surveys: A Step-by-Step Guide* (4<sup>th</sup> ed.). California: SAGE Publications, Inc.
- Finn, J. D., & Achilles, C. M. (1990). (1990). Answers and Questions about Class Size: A Statewide Experiment. *American Educational Research Journal* 27 (3), 557-577.
- Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2003). The “whys” of class size: student behaviour in small classes. *Review of Educational Research*, 73(3), 321-368.
- Fox, M. J. (1995). *Quality Assurance Management*. 2<sup>nd</sup> edition. London: Chapman & Hall.
- Freeman, R. (1993). *Quality assurance in training and education – How to apply BS5750 (ISO 9000) standards*. London: Kogan Page.
- Gardner, H. (2006). *Assessment and Learning*. Thousand Oaks, California: SAGE Publications.
- Gasemann, K. (1993). The role of Performance assessment in education. In E. Kangasniemi & S. Takala (Eds.), *Pupil assessment and the role of final examinations in secondary education*. Report of the Educational Research Workshop held in Jyvaskyla (Finland). Jyvaskyla: Swets and Zeitlinger BV Publishers, Lisse.

- Gipps, C., (1995). Reliability, Validity and Manageability in large-scale performance assessment. In Torrance, H. (Ed.), *Evaluating Authentic Assessment*. Buckingham: Open University Press.
- Glaser, B., & Strauss, A. (1967). *The discovery of Grounded Theory*. Chicago: Aldine.
- Goddard III, D., & Villanova, P. (2006). Designing Surveys and Questionnaires for Research. In Leong, F. T. L., & Austin, J. T. (Eds.), *The Psychology Research Handbook*. California: Sage Publications.
- Goetsch, D. L., & Davis, S. B. (1997). *Introduction to total Quality: Quality management for Production Processing and Services*. (2<sup>nd</sup> ed.). New Jersey: Prentice-Hall, Inc.
- Goodman, G. S., & Carey, K. T. (2004). *Ubiquitous Assessment*. New York: Peter Lang Publishes.
- Government of Botswana. (2006). *A Study to Establish the National Qualifications Framework*. Gaborone. Botswana.
- Government of Botswana. (1993). *Report of the National Commission on Education*. Gaborone: Government Printer
- Government of Botswana. (1994). *Revised National Policy on Education*. Gaborone: Government Printer.
- Government of Botswana. (1977). *Education for Kagisano: Report of the National Commission on Education*. Gaborone: Government Printer.
- Gravemeijer, K. (1998). Developmental Research as a research method. In J. Kilpatrick and A. Sierpiska (Eds.), *Mathematics Education as a Research Domain: A search for identity* (pp.277 -95). Dordrecht: Kluwer Academic Publishers.
- Gravemeijer, K. (2006). Developmental Research as a research method. In J. Van den Akker, K. Gravemeijer, S. McKenney & N. Nieveer (eds). *Educational Design Research* (pp 115 – 131). London: Routledge.

- Gravemeijer, K., & Cobb, P. (2006). Design Research from a learning design perspective. In Van den Akker, J. K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational Design Research* (pp 8 – 13). London: Routledge.
- Greaney V., & Kellagan, T. (2001) *Using Assessment to improve the quality of education*. Paris: UNESCO.
- Greenwood, M. S., & Gaunt, H. J. (1994). *Total Quality Management for Schools*. London: Cassell.
- Griffith University Research Higher Degree Handbook. (2005). Retrieved from <http://www.griffith.edu.au/>
- Grima, G. (2004). *The Secondary Education Certificate Examination in Malta: An Evaluation of the Differentiated Paper System*. A paper presented at the third Conference of the Association of Commonwealth Examination and Association Boards. Nadi, Fiji.
- Grima, G., & Ventura. F. (1998). *School-based Assessment in Malta: lessons from the past, directions for the future*. Paper presented at First ACEAB Conference. Trou aux Biches. Mauritius.
- Grima, G., & Ventura. F. (2000). *School-based Assessment in Malta: lessons from the past, directions for the future*. Paper presented at First ACEAB Conference 4-8 September. Trou aux Biches. Mauritius.
- Grisay, A., & Mahlck, L. (1991). *The quality of education in developing countries: A review of some research studies and policy documents*.
- Gronlund, N. E. (2003). *Assessment of students Achievement (7<sup>th</sup> ed.)*. Boston: Allyn and Bacon.
- Gronlund, N. E. (2006). *Assessment of students Achievement (8<sup>th</sup> ed.)*. Boston: Allyn and Bacon.
- Guba, E. G., and Lincoln, Y. S. (1981). *Effective Evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass.

- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA: SAGE.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: SAGE.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate Data Analysis with Readings*. New Jersey: Prentice Hall, Inc.
- Halsall, R. (1998). School Improvement: An overview of key findings and messages. In R. Halsall, (Ed.), *Teacher Research and school Improvement: Opening doors from the inside*. Buckingham: Open University Press.
- Hammersley, M., & Atkinson, P. (1983). *Ethnography: Principles in Practice*. London: Tavistock.
- Hamp-Lyons, L. (2009). *The impact of assessment reform on teachers' constructs of oral interaction in English in Hong Kong*. A paper presented at the 35<sup>th</sup> IAEA Conference. Australia.
- Hargreaves, E. (2007). The validity of collaborative assessment for learning. *Assessment in Education*. 14(2), 185-199.
- Harlen, W., Gipps, C., Broadfoot, P., & Nuttal, D. (1992). Assessment and the improvement of education. *The Curriculum Journal*, 3(3), 215-30.
- Harlen, W. (1994). *Enhancing Quality in Assessment*. London: Paul Chapman Publishing Ltd.
- Harlen, W., & Deaken-Crick, R. (2002). "A systematic review of the impact of summative assessment and tests on students' motivation for learning (EPPI-Centre Review, version 1.1)". In *Research Evidence in Education Library*, issue 1. EPPI-Centre, Social Science Research Unit, Institute of Education. London.
- Harlen, W. (2006). *Teaching, learning & assessing science 5-12*. London: SAGE Publications Ltd.

- Haynes, A. B. (2000). *Current practices and future possibilities for the Caribbean Examinations Council's (CXC) School-Based Assessment*. Paper presented at First ACEAB Conference 4-8 September. Trou aux Biches. Mauritius.
- Henderson, G. R. (2006). *Six Sigma: Quality Improvement with MINITAB*. West Sussex: John Wiley & Sons, Ltd.
- Herman, J. L., Baker, E.L., & Linn, R. L. (2006). "Assessment for accountability and learning", *CRESST LINE, Newsletter of the National Center for Research on Evaluation, Standards, and Student Testing*, Fall edition.
- Hoskin, K. (1979). The examination, disciplinary power and rational schooling. In *History of education*. (8), 135-146. London: Taylor & Francis.
- Hoadley, C. P. (2002). *Creating context: Design-based research in creating and understanding CSCL. Proceedings of Computer Support for Cooperative Learning (CSCL)*. Bourder, Co.
- Hornsby, A.S. (2000). *Oxford Advanced Learners' Dictionary*. Sixth Edition. Oxford: University Press.
- Howie, S. J., & Plomp, T. (2001). *English Language Proficiency and Other Factors Influencing Mathematics Achievement at Junior Secondary Level in South Africa*. A paper presented at the annual meeting of the American Educational Research Association. Seattle, WA.
- Howie, S. J. (2002). *English language proficiency and contextual factors influencing mathematics achievement of secondary school pupils in South Africa*. Doctoral dissertation. Enschede (NL): Print Partners Ipskamp.
- Howie, S. J., & Plomp, T. (2003). *Language Proficiency and Contextual Factors Influencing Secondary Students' Performance in Mathematics in South Africa*. A paper presented at the annual meeting of the American Educational Research Association. Chicago. April 21-25.



- Howie, S. J. (2006). *Assessment for Learning Strategies: merging theory and practice within South Africa realities*. A paper presented at the GDE Assessment for Learning Conference. Johannesburg.
- Hoxby, C. M. (2000). The Effects of Class Size on Student Achievement: New Evidence from Population Variation. *Quarterly Journal of Economics*, 115(4): 1239-1285.
- Islam, K. A. (2006). *Designing and measuring: The 6 Sigma way*. San Francisco: Pfeiffer, An Imprint of Willey.
- Izard, J. F. (1998). Validating teacher-friendly (and student-friendly) assessment approaches. In D. Greaves & P. Jeffery (Eds.) *Strategies for intervention with special needs students*. (pp.101-115). Melbourne, Vic.: Australian Resource Educators' Association Inc..
- Izard, J. F. (2002). *Describing student achievement in teacher-friendly ways: Implications for formative and summative assessment*. Valetta, Malta: Ministry of Education, Malta and the University of Malta for the Association of Commonwealth Examinations and Accreditation Bodies.
- Januario, F. M. (2008). *Investigating and improving assessment practices in physics in secondary schools in Mozambique*. Doctoral dissertation. University of Pretoria.
- James, M. (1994). Experience of quality assurance at key stage 1. In W. Harlen (Ed.), *Enhancing Quality in Assessment*. London: Paul Chapman Publishing Ltd.
- James, M. (2006). Assessment, Teaching and Theories of Learning. In Gardner, J (Ed.), *Assessment and Learning*. London: SAGE Publications.
- Jewel, T., & Ford, M. (2007). What exactly do you want me to do? Analysis of a Criterion Referenced Assessment Project. *Journal of Information Technology Education*. Online. Retrieved 22 May 2008. Vol 6. 311-326.
- Johnson, R., Penny. J. A., & Gordon., B. (2009). *Assessing Performance – Designing, Scoring and Validating Performance tasks*. New York: The Guilford Press.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.

- Jones, D. P. (2002). *Different Perspectives on Information about Educational Quality: Implications for the Role of Quality*. Washington, D. C: Council for Higher Education Accreditation.
- Jones, V. (2006). How do teachers learn to be effective Classroom managers?. In C. M. Evertson, & C. S. Weinstein, (Eds.), *Handbook of classroom Management: Research, Practice, and Contemporary Issues*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Jordan, P., & McDonald, J. (2008). Data presented in a *symposium "How standards of student achievement work to support teacher judgement: The place of moderation"*. Australian Association for Research in Education Conference, Brisbane.
- Kanjee, A., & Y. Sayed, Y. (2008). *Assessment and Education Quality in South Africa*. A paper presented at the world 52nd Annual Meeting of the Comparative and International Education Society. Teachers College, Columbia University New York, 17-21 March.
- Kane, M. T. (2008). Terminology, Emphasis, and Utility in Validation. *Educational Researcher*. 37(2),76-82.
- Kapur, K. (2008). *Assessment for Improving Learning in Schools in India : A perspective*. A paper presented at the IAEA Annual Conference. Cambridge. UK.
- Kelly, A. (2004). Design Research in Education: Yes, but is it Methodological? *Journal of the Learning Sciences*. 13(1), 115-128.
- Keightley, J. V., & Coleman, M. J. (2002). *Improving the Quality of Education using School-Based Assessment: Advantages, Disadvantages, Issues and Challenges*. Paper presented to the International Association for Educational Assessment. UNESCO sponsored Round Table. Hong Kong.
- Keightley, J. V. (2002). *School-based Assessment in South Australia*. Paper presented to Hong Kong Examinations and Assessment Authority Seminar. Latest Developments in Educational Assessment.

- Kellagan, T., & Greaney, V. (2001) *Using Assessment to improve the quality of education*. Paris: UNESCO.
- Kellaghan, T., & Greaney, V. (2003). *Monitoring Performance: Assessment and Examinations in Africa training*.
- Kent, T. W., & McNergney, R. F. (1999). *Will technology really change education?* Thousands Oaks, CA: Corwin Press.
- Knostantopolous, S. (2008). Do Small Classes Reduce the Achievement Gap between Low and High Achievers? Evidence from Project STAR. *Elementary School Journal*, 108(4), 275-291.
- Knostantopolous, S., & Chung, V. (2009). What are the Long Term Effects of Small Class on the Achievement Gap? Evidence from the Lasting Benefits Study. *American Journal of Education*, 116(1), 125-154.
- Khoo. H. C. S., & Idrus, R. M. (2004). *A Study of Quality Assurance Practices in the Universiti Sains Malaysia (USM)*. . Turkish Online Journal of Distance Education. 5(1). Retrieved 05 July 2009.
- Klenowski, V., & Wyatt-Smith, C. (2008). “Standards-driven reform Years 1–10: Moderation an optional extra?”, A paper presented at the Australian Association for Research in Education Conference, Brisbane.
- Kobrin, J., & Kimmel, E. (2006). *Test development and technical information on the writing section of the SAT reasoning test*. New York: The College Board. Retrieved January 5, 2009, from [www.collegeboard.com/research/pdf/pdf/RN-25.pdf](http://www.collegeboard.com/research/pdf/pdf/RN-25.pdf).
- Kremelberg, D. (2011). *Practical Statistics*. California: SAGE Publications, Inc.
- Ladson-Billings, G., & Donnor, J. (2005). The moral activist role of critical race theory scholarship. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (3<sup>rd</sup> ed, pp.279-301). Thousand Oaks, CA: Sage.
- Lane, S., & Stone, C. (2006). Performance assessment. In R. Brennan (Ed.), *Educational measurement*. Westport, CT: American Council on Education and Praeger.

- Lather, P. (1992). Critical frames in educational research: Feminist and post-structural perspectives. *Theory and practice*, 31(2), 1-13.
- Lee, R. M., & Fielding, N. G. (2009). Tools for qualitative data analysis. In ? Hardy & Bryman (Eds.), *The handbook of data analysis*. London: Sage Publications Ltd.
- Le Grange, L., & Reddy, C. (1998). *Continuous Assessment: An introduction and Guidelines to Implementation*. Cape Town: Juta & Company Limited.
- Lennox, B. (2000). *Achieving National Consistency in School-Based Assessment against Standards*. Paper presented at First ACEAB Conference. Trou aux Biches, Mauritius.
- Lincoln, Y. S., & Guba, E. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.
- Lincoln, Y. S., & Guba, E. (1999). Establishing Trustworthiness. In Bryman, A. & Burgess, R. G. (eds.), *Qualitative Research Volume III*. London: SAGE Publications.
- Linn, R. L. (2000) "Assessments and accountability", *Educational Researcher*, 29, (2), 4–16.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1993). Policy and Validity prospects for performance –based assessment. *American Psychologist*, 48(12), 10-18.
- Loken, R. D. (1973). *Education in Transition: The report of the Polytechnic Mission*.
- Madaus, G., & O'Dwyer, L. (1999). A short History of Performance assessment. *Phi Delta Kappan*, 80(9), 688-695.
- Mafumiko, F. (2006). *Micro-scale experimentation as a catalyst for improving the chemistry curriculum in Tanzania*. Doctoral dissertation. Enschede (NL): University of Twente.
- Mamary, A. (2007). *Creating the ideal school – where teachers want to teach and students want to learn*. Plymouth: Rowman & Littlefield Education.
- Masole, T., M. (2003). *Teachers' perception about assessment and contribution of BGCSE Agriculture continuous assessment towards final grade*. Unpublished Master of Education Dissertation. University of Botswana.
- Marzarno, in Nitko & Brookhart (2007). *Educational Assessment of Students*. New Jersey: Merrill Prentice Hall.

- Masters, G. N., & McBryde, B. (1994). *An Investigation of the Comparability of Teachers' Assessments of Student Folios*. Tertiary Entrance Procedures Authority. Brisbane.
- Maughan, S. (2004). *Closing the gap between assessment and learning*. A paper presented at the third Conference of the Association of Commonwealth Examination and Association Boards. Nadi, Fiji.
- Maxcy, S. J. (2003). Pragmatic threads in mixed methods research in the social sciences: The search for multiple modes of inquiry and the end of philosophy of formalism. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioural research*. (pp 51-90). Thousand Oaks, CA: Sage.
- Maxwell, G. S. (2004). *Progressive assessment for learning and certification: Some lessons from School-based assessment in Queensland*. A paper presented at the third Conference of the Association of Commonwealth Examination and Association Boards. Nadi, Fiji.
- May, D. (2006). *Geography of Botswana*. Gaborone: McMillan.
- McDaniel, C., & Gates, R. (2010). *Marketing Research with SPSS*. (8<sup>th</sup> ed.). Danvers MA: John Wiley & Sons, Inc.
- McDavid, J. C., & Hawthorn, L. R. L. (2006). *Programme Evaluation & Performance Measurement: An Introduction to Practice*. California: SAGE Publications, Inc.
- McIntire, S. A., & Miller, L. A. (2007). *Foundations of Psychological Testing: A practical Approach*. (2<sup>nd</sup> ed.). London: Sage Publications, Inc.
- McKenny, S. (2001). *Computer-based support for science education material developers in Africa: exploring potentials*. Doctoral dissertation. Enschede: University of Twente.
- McMillan, J. H. (2004). *Classroom Assessment: Principles and Practice for Effective Instruction*. Boston: Allyn and Bacon.
- McMillan, J. H. (2000). Fundamental Assessment Principles for Teachers and School Administrators. *Practical Assessment, Research and Evaluation: A Peer-reviewed electronic journal*, 7 (8).

- Mehrens, W. A., & Lehman, I. J. (1991). *Measurement and Evaluation in Education and Psychology*. Florida: Holt, Rinehart and Winston, Inc.
- Mehrens, W. A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice*, 11(1), 3-9.
- Mergendoller, J. R., Markham, T. Ravitz, J., & Larmer, J. (2006). Pervasive management of Project Based Learning: Teachers as guides and facilitators. In Evertson, C. M. & Weinstein, C. S. (Eds.), *Handbook of classroom Management: Research, Practice, and Contemporary Issues*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Mertens, D. M. (2005). *Research and evaluation in education and psychology: Integrating diversity with qualitative, quantitative and mixed methods* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- Mertens, D. M. (2010). *Research and Evaluation in Education and Psychology*. California: SAGE Publication, Inc.
- Mertler, C. A. (2001). *Designing Scoring Rubrics for your classroom. Practical Assessment, Research & Evaluation*. 2(3). Retrieved February 14, 2005 from <http://pareonline.net/getvn.asp?v=7&n=25>.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement*. New York: American council on Education and Macmillan.
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied Multivariate Research: Design and Intepretation*. California: Sage Publications.
- Mehrens, W. A. (1992). Using Performance assessment for accountability purposes. *Educational Measurement: Issues and practice*, 11(1), 3-9.
- Mercurio, A. (2008). *Re-imagining school-based assessment at upper education level*. A paper presented at the A paper presented at IAEA conference, Cambridge.
- Milesi, C., & Gamoran, A. (2006). Effects of Class Size and Instruction on Kindergarten Achievement. *Educational Evaluation and Policy Analysis*, 28(4), 287-313.

- Miller-Jones, D. (1989). Culture and Testing. *American Psychologist*, 44, 360-366.
- Miller, D. C., Sen, A., & Malley, L. B. (2007). *Comparative Indicators of Education in the United States and Other G-8 Countries: 2006*. (NCES 2007-006). National Centre for Education Statistics, Institute of Education Sciences, U. S. Department of Education. Washington, D. C.
- Mindes, G. (2007). *Assessing Young Children*. New Jersey: Merrill Prentice Hall
- Ministry of Education & Skills Development. (2001). *Botswana General Certificate of Secondary Education Assessment Syllabus: Agriculture*. Gaborone: Government Printer.
- Ministry of Education & Skills Development. (2000a). *Botswana General Certificate of Secondary Education Agriculture Teaching Syllabus*. Gaborone: Government Printer.
- Ministry of Education & Skills Development. (2000b). *Botswana General Certificate of Secondary Education Agriculture Assessment Syllabus*. Gaborone: Government Printer.
- Ministry of Education & Skills Development. (2001). *Early Childhood Care and Education Policy*. Gaborone: Government Printer.
- Ministry of Education & Skills Development. (2002a). *Lower Primary School Syllabus: Standard One to Four*. Gaborone: Government Printer.
- Ministry of Education & Skills Development. (2002b). *Curriculum Blueprint: Senior Secondary Education Programme*. Gaborone: Government Printer.
- Ministry of Education & Skills Development. (2002c). *Curriculum Blueprint: Ten-Year Basic Education Programme*. Gaborone: Government Printer.
- Ministry of Education & Skills Development. (2005). *Upper Primary School Syllabus: Standard Five to Seven*. Gaborone: Government Printer.

- Ministry of Education & Skills Development. (2006). *Trends in International Mathematics and Science Study- 2003*. Gaborone: Government Printer.
- Ministry of Education & Skills Development. (2006). *Education Statistics 2004*. Gaborone: Government Printer.
- Ministry of Education & Skills Development. (2007). *Curriculum Blueprint: The Ten-Year Basic Education Programme*. Gaborone: Government Printer.
- Ministry of Education & Skills Development. (2009). *Standard Four Assessment Project*. Gaborone: Government Printer.
- Ministry of Education & Skills Development. (2009a). *Trends in International Mathematics and Science Study-2007*. Gaborone: Government Printer
- Ministry of Education & Skills Development. (2009). *Education Statistics 2006*. Gaborone: Government Printer.
- Ministry of Finance and Development Planning. (1991). *National Development Plan 7, 1991-1997 (NDP 7)*. Gaborone: Government Printer.
- Ministry of Finance and Development Planning. (1997). *National Development Plan VIII (1997 - 2003)*. Gaborone: Government Printer
- Ministry of Finance and Development Planning. (2001). *Population and Housing Census*. Gaborone: Government Printer.
- Ministry of Finance and Development Planning. (2003). *National Development Plan 9, 2003/04-2008/09*. Gaborone: Government Printer.
- Ministry of Finance and Development Planning. (2005). *Population Projections for Botswana 2001-2031*. Gaborone: Government Printer.
- Ministry of Finance and Development Planning. (2006). *Budget Speech*. Gaborone Government: Printer.
- Ministry of Finance and Development Planning. (2009). *Budget Speech*. Gaborone: Government Printer.



- Ministry of Trade and Industry. (2007). *Botswana Review*. 27<sup>th</sup> Edition.
- Ministry of Trade and Industry. (2008). *Botswana Review*. 28<sup>th</sup> Edition.
- Morgan, D. L. (2007). Paradigm lost and pragmatism regained: methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1, 48-76.
- Moskal, B. M., & Leydens. J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10). Retrieved February 14, 2005 from <http://pareonline.net/getvn.asp?v=7&n=10>.
- Motswiri, M. J. (2004). *Supporting chemistry teachers in implementing formative assessment of investigative practical work in Botswana*. Doctoral dissertation. Enschede (NL): Print Partners Ipskamp.
- Nam, C. S., & Smith-Jackson, T. L. (2007). Web-Based Learning Environment: A Theory-Based Design Process for Development and Evaluation. *Journal of Information Technology Education*. 6, 25-43.
- Neill, M. D., & Median, J. N. (1992). Eliminating Standardised tests would improve schools. In Cozic, C. P (ed) *Education in America*. San Diego: Greenhaven Press.
- Nenty, H. J., Odili, J. N., & Munene-Kabanya, A. N. (2008). *Assessment training among secondary school teachers in Delta state of Nigeria: implications for sustaining standards in Educational Assessment*. A paper presented at the 28<sup>th</sup> AEAA, Conference. Accra. Ghana.
- Nichols, P. D., & Williams, N. (2009). Consequences of Test Score use as Validity Evidence: Roles and Responsibilities. *Educational Measurement: Issues and Practice*, 28(1), 3-9.
- Nir-Gal, O. & Klein, P. S. (2004). Computers for cognitive development in Early Childhood - The Teacher's role in the computer Learning Environment. *Information Technology in childhood Education Annual*, 4(1), 97-119.
- Nitko, A. J. (1995). Curriculum-based Continuous assessment: a frame work for concepts, policies and procedures. *Assessment in Education*, 2,321-337.

- Nitko, A. J. (1998). *The Role of Examination Research and Testing Division in Continuous Assessment Reform in Botswana*. Examination Research and Testing Division. Gaborone.
- Nitko, A. J. (2004). *Educational Assessment of Students*. New Jersey: Merrill Prentice Hall.
- Nitko, A. J., & Brookhart, S. M. (2007). *Educational Assessment of Students*. New Jersey: Merrill Prentice Hall.
- Njabili, A. F. (1987). *Continuous Assessment: The Tanzanian experience*. In P. Broadfoot, H. Torrance, & R. Murphy, (Eds.), *Changing Educational Assessment: International Perspectives and trends*. London: Routledge and Kegan Paul.
- Neuman, W. L. (2000). *Social Research Methods: Qualitative and Quantitative Approaches*. Boston: Allyn and Bacon.
- Noor, A. G. (2008). *Oral and Practical testing in public examinations (Kenyan experience): Challenges and prospects*. A paper presented at the 28<sup>th</sup> AEAA, Conference. Accra. Ghana.
- Nowa-Phiri, M. D. (2000) *Performance Assessment: is it possible in a country where cheating in National Examinations is rampant?*. Paper presented at the First ACEAB Conference 4-8 September. Trouaux Biches, Mauritius.
- Nye, B., Hedges, L. V., & Konstantopoulos, S. (2002). Do Low-Achieving Students Benefit More from Small Classes? Evidence from the Tennessee Class Size Experiment. *Educational Evaluation and Policy Analysis* 24(3), 201-217.
- Oakland, J. S. (2003). *Statistical Process Control*. Oxford: Butterworth Heinemann.
- Onyango, P. O., & Ndege, J. G. (2007). Linking school-based assessment with public examinations: The Kenya National Examinations Council Experience. *Journal of Educational Assessment in Africa*. 1(1)24-32.
- Ornstein, A. C., & Hunkins, F. (1993). *Curriculum: Foundations, Principles and Theory*. Boston: Allyn and Bacon.

- Ottevanger, W. J. W. (2001). *Teacher support materials as a catalyst for science curriculum implementation in Namibia*. Doctoral dissertation. Enschede (NL): Print Partners Ipskamp.
- Patchen, M. (2004). *Making our schools more effective: what matters and what works*. Illinois: Charles C Thomas Publisher, LTD.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (2<sup>nd</sup> ed.), Thousand Oaks, CA: SAGE.
- Pelliegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing What Students Know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Persse, J. R. (2006). *Process Improvement Essentials*. Sebastopol: O'Reilley Media, Inc.
- Pearson, R. W. (2010). *Statistical Persuasion*. Thousand Oaks, California: SAGE Publications Inc.
- Piaget, J. (1953). *The origin of intelligence in the child*. London: Routledge and Kegan Paul.
- Pitman, J. (2003). Preparing Teachers to Use Technology with Young Children in Classroom. *Information Technology in Childhood Education Annual*, 4(1), 261-287.
- Plomp, T. (2008). *Educational Design Research: An Introduction*. A paper presented at the workshop of the SANPAD Project. University of Pretoria.
- Plumer, R. (1990). In. D. Lock & D. J. Smith (Eds.), *Gower handbook of Quality management*. Gower Publishing Company Limited. Aldershot.
- Pong, S., & Pallas, A. (2001). Class Size and Eighth-Grade Math Achievement in the United States and Abroad. *Educational Evaluation and Policy Analysis*, 23 (3), 251-273.
- Pongi, V. (2004). *Making the switch from "Assessment for ranking" towards "Assessment for Learning"; The challenges facing the small Islands States of the Pacific*.
- Popham, W. J. (2005). *Classroom Assessment: What teachers need to know*. Boston: Pearson Education, Inc.

- Portal, M. (2000). *School Based Assessment: Problems and Solutions*. Paper presented at The First ACEAB Conference. Trou aux Biches. Mauritius.
- Queensland Studies Authority. (1998). *Strategies for authenticating student work for learning and assessment*. Queensland Government.
- Queensland Studies Authority. (2008). *The Quality Assurance of Authority-registered Subjects*. Queensland Government.
- Queensland Studies Authority. (2009). *Student assessment regimes: getting the balance right for Australia*. Queensland Government.
- Radnor, H., & Shaw, K. (1995). 'Developing a Collaborative Approach to Moderation. In H. Torrance, (Ed.), *Evaluating Authentic Assessment*. Buckingham: Open University Press.
- Raffan, J. (2000). *School-Based Assessment: Principles and Practice*. Paper presented at the First ACEAB Conference. Trou aux Biches. Mauritius.
- Rainey, D. (2005). *Product Innovation: Leading change through integrated product development*. New York: Cambridge University Press.
- Raivoce & Pongi, (2000). *Performance Assessment at the Pacific Senior Secondary Certificate (PSSC): the SPBEA Experience*. Paper presented at the First ACEAB Conference. Trouaux Biches, Mauritius.
- Rennert-Ariev, P. (2005). Theoretical model for the authentic assessment of teaching, Practical Assessment, Research and Evaluation. *A Peer-reviewed electronic journal*. 10 (2).
- Republic of Botswana. (1977). *Education for Kagisano: Report of the National Commission on Education*. Gaborone: Government Printer.
- Republic of Botswana. (1993). *Report of the National Commission on Education*. Gaborone: Government Printer.
- Republic of Botswana. (1994). *Revised National Policy on Education*. Gaborone: Government Printer.

- Republic of Botswana (2008). *Public Service Act No 30 of 2008*. Gaborone: Government Printer.
- Richard, F. (1993). *Quality assurance in training and education – How to apply BS5750 (ISO 9000) standards*. London: Kogan Page.
- Rudner, L. M. & Boston, C., (1994). Performance Assessment, *ERIC Review*, 3(1) 2-12.
- Republic of Botswana. (2009). *Preliminary Botswana HIV/AIDS Impact Survey III Results*. Gaborone: Government Printer.
- Rudner, L. M., & Boston, C. (1994). Performance assessment, *ERIC Review*, 3(1), 2-12.
- Ryan, T. (2006). Performance assessment. Critics, criticism, and controversy. *International Journal of Testing*, 6(1), 97-104.
- Salvia, J., & Ysseldyke, J.E. (1998). *Assessment*. Boston: Houghton Mifflin Company.
- Schostak, J. (2006). *Interviewing and Representation in Qualitative Research*. Open Glasgow: University Press.
- Schwandt, T. A. (2000). Three epistemological stances for qualitative inquiry; interpretivism, hermeneutics, and social constructionism. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2<sup>nd</sup> ed., pp 189-214). Thousand Oaks, CA: SAGE.
- Seale, C. (2004). *Social Research Methods*. London: Routledge.
- Shaklee, B. D., Barbour, N. E., Ambrose, R., & Hansford, S. J., (1997). *Designing and using portfolios*. Boston: Allyn and Bacon.
- Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the Science of Education Design Studies. *Educational Researcher*. 32 (1), 25-28.
- Shepherd, L. A. (2008). “A brief history of accountability testing 1965–2007”. In K. Ryan & L. Shepard (Eds.), *The Future of Test-Based Educational Accountability*, New York: Routledge.
- Shepherd, L.A. (2000). “The role of assessment in a learning culture”, *Educational Researcher*.

- Singh, P. (2000). *Implementing School-Based Assessment; A Functional Approach*. Paper presented at the First ACEAB Conference. Trouaux Biches. Mauritius.
- Singh, T. (2004). *School-Based Assessment: the interface between Continuous Assessment (CASS) and the external summative examinations at Grade 12 level with special focus on Mathematics and Science*. Unpublished Master of Education dissertation. University of Pretoria: Pretoria.
- Slavin, R. (1994). *Educational Psychology: Theory and Practice*. Boston: Allyn and Bacon.
- Stanley, G., & Tognolini, J. (2008). “*Performance with respect to standards in public examinations*”. A paper presented at the Annual Conference of the International Association for Educational Assessment. Cambridge.
- Stevens, J. P. (2002). *Applied Multivariate Statistics for the Social Sciences*. New Jersey Mahwah: Lawrence Erlbaum Associates Publishers.
- Stiggins, R. J. (1997). *Student-Centred Classroom Assessment* (2<sup>nd</sup> ed.). New Jersey: Prentice-Hall, Inc.
- Stiggins, R. J. (2002). Where is our assessment future and how can we get there from here? In Lissitz & Schafer, (Eds.), *Assessment in Educational Reform: Both means and ends*. Boston: Allyn and Bacon.
- Stiggins, R. J. (2002). Assessment Crisis: The absence Of Assessment *FOR* Learning. *Phi Delta Kappan*, 83(10), pages 758 – 765. Retrieved, 03 March, 2009.
- Stiggins, R. J. (2009) “Assessment, student confidence, and school success”, *Phi Delta Kappan*.
- Stobart, G. (2008). *The validity of ability tests - a case of over-interpretation?* A paper presented at the 34<sup>th</sup> IAEA Conference. London.
- Tabachnik, B. G., & Fidell, L. S. (2001). *Using Multivariate statistics*. Needham Heights, MA: Allyn & Bacon.
- Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social & behavioural research*. Thousand Oaks, CA: Sage.

- Tecele, A. T. (2006). *The potential of a professional development scenario for supporting biology teachers in Eritrea*. Doctoral dissertation. Enschede (NL): University of Twente.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research*. Thousand Oaks, CA: Sage).
- The Design-Based Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher*, 32(1), 5-8.
- Thomas, T., Davis, T. & Kazlauskas, A. (2007). Embedding critical thinking in IS Curricula. *Journal of Information Technology Education*. 6, online, 327-346.
- Tlou, T., & Campbell, A. (1984). *History of Botswana*. Gaborone: McMillan.
- Torrance, H., & Pryor, J. (1998). *Investigating formative Assessment: Teaching, Learning and Assessment in the classroom*. Buckingham: Open University Press.
- Thorndike, R. M., & Thorndike-Christ, T. (2010). *Measurement and Evaluation in Psychology and education*. New York: Pearson Education, Inc.
- Tilya, F. N. (2003). *Teacher support for the use of MBL in Activity-Based Physics teaching in Tanzania*. Doctoral dissertation. Enschede (NL): University of Twente.
- Tindal, G. & Haladyna, T. M. (2002). *Large-Scale Assessment Programs for all students – Validity, Technical Adequacy and Implementation*. New Jersey: Lawrence Erlbaum Associates Publishers, Inc.
- Torrance, H. (1995). *Evaluating Authentic Assessment*. Buckingham: Open University Press.
- Torrance, H., & Pryor, J. (1998). *Investigating formative Assessment: Teaching, Learning and Assessment in the classroom*. Buckingham: Open University Press.
- UNESCO. (1990). *World Declaration on Education For All. Meeting basic learning needs*. New York.
- UNESCO. (2000a). *The Dakar Framework for Action: Education for All – Meeting our Collective Commitments*. World Education Forum, Dakar, Senegal. 26 – 28 April. Paris: UNESCO.

- UNESCO. (2000). *The Dakar Framework for Action: Education for All – Meeting our Collective Commitments*. World Education Forum, Dakar, Senegal. 26–28 April. Paris: UNESCO.
- UNESCO. (2002). *Education For All. Is the World on track? EFA Global Monitoring Report*. Paris: UNESCO.
- UNICEF. (1990). *Meeting Basic Learning Needs: A Vision for the 1990's World Conference on Education For All*. 5-8 March. Jomtein. Thailand.
- United Nations. (2000). *United Nations Millennium Declaration. Resolution adopted by the General Assembly*. (United Nations A/RES/5/2) [www.un.org/millennium/declaration/ares552e.htm](http://www.un.org/millennium/declaration/ares552e.htm)
- United Nations. (2001a). *Committee on the Rights of the Child, General Comment 1: The Aims of education*. Washington. D. C.
- United Nations. (2005). *The Millennium Development Goals Report*. New York.
- Van den Akker, J., & Plomp, T. (1993). *Development Research in Curriculum: Propositions and experiences*. A Paper presented at AERA Annual Meeting, April 12-16, Atlanta.
- Van den Akker, J. K. (1999). Principles and Methods of Development Research. In J. Van den Akker, R. M. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds.), *Design approaches and tools in education and training*. Boston: Kluwer Academic, 1-14.
- Van den Akker, J., Branch, R., Gustafson, K., Nieveen, N., & Plomp, T. (eds) (1999). *Design Approaches and Tools in Education and Training*. Dordrecht: Kluwer Academic Publishers.
- Van den Akker, J. K., Gravemeijer, S., McKenney, S. & N. Nieveen. (2006). *Educational Design Research* (pp 8 – 13). London: Routledge.
- Van der Berg, S., & Shepherd, D. (2010). *Signalling performance: continuous assessment and matriculation examination marks in South African schools*. Stellenbosch Economic Working Papers: 28/10 Stellenbosch University. Bureau for Economic Research



- Van der Merwe, I. F. J. (2000). *Continuous Assessment: The Namibian Experience*. Paper presented at the First ACEAB Conference. 04–08 September. Trou aux Biches, Mauritius.
- Viswanathan, M. (2005). *Measurement Error and Research Design*. Thousand Oaks, California: SAGE Publications.
- Vygotsky, L. (1978). *Mind in society. The development of higher mental processes*. Cambridge, Mass: Harvard University Press.
- Walker, D. (2006). *Toward Productive Design Studies*. In J. Van den Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational Design Research* (pp 8–13). London: Routledge.
- Walklin, L. (1992). *Putting quality into practice*. Cheltenham: Stanley Thorns (Publishers). Ltd.
- Wiersma, W., & Jurs, S. G. (2005). *Research Methods in Education: An Introduction*. Boston: Pearson Education, Inc:
- Wiggins, G. (1998). *Educative Assessment – Designing Assessments to Inform and Improve Student Performance*. San Francisco: Jossey-Bass Publishers.
- Wiggins, S., & Riley, S. (2010). QMI: Discourse Analysis. In Forester (Ed.), *Doing Qualitative Research in Psychology: A Practical Guide*. New Delhi: SAGE.
- Wild, C. L., & Ramaswamy, R. (2008). *Improving Testing: Applying Process Tools and Techniques to Quality*. New York: Lawrence Erlbaum Associates.
- Wiles, J. & Bondi, J. (2000). *Supervision: A guide to practice* (5<sup>th</sup> ed.). New Jersey: Prentice Hall, Inc.,
- William, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment. *British educational Research Journal*, 22(3), 537-48.
- Willig, C. (2001). *Introducing Qualitative Research in Psychology: Adventures in Theory and Method*. Buckingham: Open University Press.

- Wood, R. (1991). *Assessment and Testing*. Cambridge: Cambridge University Press.
- Yadidi, D.C., & Banda, A. C. (2008). Making assessment for the promotion of teaching and learning in school-the case of the Malawi School Certificate of Education Exams (MSCE). *Journal of Educational Assessment in Africa*. 2(1), 44-52.
- Yandila, C. D., Komane, S. S., & Moganane, S. (2003). *Evidence of hands-on Teaching/Learning Approaches in Botswana's Senior Secondary school Science Lessons*. A paper presented at World Conference on Science and Technology Education. 7-10 April. Penang, Malaysia.
- Yao, Y., Thomas, M., Nickens, N., Downing, J. A., Burkett, R., S., & Lamson, S. (2008). Validity Evidence of an Electronic Portfolio for Pre-service teachers. *Educational Measurement: Issues and Practice*. 27(1), 10-24.
- Yin, R. K. (2009). *Case study research: Designs and methods* (4<sup>th</sup> ed.). Thousand Oaks, CA: Sage.

## APPENDICES

See attached CD.

- a. Appendix 2.1: Grade descriptors
- b. Appendix 2.2: Examples of tasks for practical
- c. Appendix 2.3: Criteria for assessing practical tests
- d. Appendix 2.4: Marksheet for scoring the project
- e. Appendix 4.1: A matrix illustrating the constructs and questions
- f. Appendix 4. 2: Teacher questionnaire
- g. Appendix 4.3: School administration questionnaire
- h. Appendix 4.4: Teachers interview schedule
- i. Appendix 4.5: Observation Schedule
- j. Appendix 4.6: Students focus group interview
- k. Appendix 4.7: Teacher Interview Guide
- l. Appendix 4.8: Student questionnaire
- m. Appendix 4.9: Teacher questionnaire
- n. Appendix 4.10: Student rerecord book
- o. Appendix 4.11: Ethics Approval
- p. Appendix 4.12: Permission from Ministry
- q. Appendix 4.13: Permission from Regions
- r. Appendix 4.14: Permission from Schools
- s. Appendix 4.15: Student Consent

- t. Appendix 4.15: Participant consent
- u. Appendix 4.16: Minor's parental consent
- v. Appendix 6.2: Skills equating for Task 2 and 3
- w. Appendices 6.3: Task 1 development
- x. Appendix 6.4: Task 3 development
- y. Appendix 6.5: Experts demographic information
- z. Appendix 6.6: Expert Evaluation
- aa. Appendix 6.7: Number of items and reliability coefficients
- bb. Appendix: 6.9: Teacher Interview schedule
- cc. Appendix 7.2: Audio recorded interviews
- dd. Appendix: 7.3 Clearance certificate