

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 teachers 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 T₈ 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 ₁	T ₂	T ₃	T ₄	T ₅	T ₆	T ₇	T ₉
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 ₁					
	T ₂					
	T ₃					

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 ₄						
	T ₅						
	T ₆						

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 ₇						
	T ₉						

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 completed 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 (Burris & 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 T₂ exhibited low understanding for assessment of learning.

Table 7.3: *The extent of teachers' embraced assessment for learning*

Instructional activities		To some extent	To a moderate extent	To a great extent	To a very great extent
1.	The teacher distributes the task to the class before the start of the practicals				T ₅ T ₇ T ₆ T ₁ T ₂ T ₃ T ₄ T ₈
2.	The teacher states the performance assessment objective before the start of the practicals			T ₆ T ₁ T ₂ T ₃ T ₄	T ₅ T ₈ T ₇
3.	The teacher asks questions to gauge the students level of knowledge of the activity		T ₄	T ₂ T ₁ T ₃ T ₅ T ₇ T ₆ T ₈	
4.	The teacher links the relevance of the practical to everyday lives		T ₄ T ₂	T ₁ T ₃ T ₈ T ₇ T ₆	T ₅
5.	The teacher clarifies what resources are to be used and how they are to be used			T ₁ T ₆ T ₂ T ₄	T ₅ T ₇ T ₃ T ₈
6.	The teacher spells out observable aspects of the student's performance/product that should be judged	T ₂		T ₁ T ₃ T ₈ T ₇ T ₄	T ₅ T ₆
7.	Teacher clarifies what will be assessed and how	T ₂		T ₁ T ₃ T ₈	T ₅ T ₇ T ₆ T ₄
8.	The teacher stresses observation of safety	T ₂	T ₁	T ₄ T ₅ T ₈ T ₆	T ₇ T ₃
9.	The teacher states the students' task			T ₂ T ₃ T ₆	T ₅ T ₇ T ₁ T ₄ T ₈
10.	The teacher states the teacher's role	T ₂		T ₆ T ₃ T ₄	T ₅ T ₇ T ₁ T ₈

11. The teachers emphasises reasoning as opposed to rote learning	T ₃	T ₅ T ₆ T ₂ T ₄	T ₈ T ₇	T ₁
12. The teacher organises material for the practical			T ₁ T ₃ T ₄ T ₅ T ₇ T ₆ T ₈	T ₂
13. The teacher manages time well		T ₃	T ₁ T ₅ T ₇ T ₆ T ₄ T ₈	T ₂
14. The teacher provides an appropriate setting to elicit and judge the performance or product			T ₂ T ₁ T ₄ T ₆	T ₃ T ₅ T ₈ T ₇
15. The teacher provides a judgement or score to describe performance		T ₁	T ₂ T ₄ T ₅ T ₈ T ₇ T ₆	T ₃

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 T₁ to T₈. 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 ₃		T ₅ T ₂ T ₄ T ₃ T ₇	T ₁ T ₈
The teacher assesses processes			T ₅ T ₂ T ₄ T ₃ T ₇	T ₃ T ₁ T ₈
The teacher forms groups during practicals			T ₅ T ₈ T ₇ T ₄	T ₃ T ₂ T ₁ T ₃
The teacher assesses individuals in group work	T ₅ T ₁ T ₇			T ₃ T ₂ T ₈ T ₄ T ₃
The teacher assesses all students in a class on the same skill in one day	T ₁ T ₄ T ₃ T ₇		T ₅ T ₂ T ₈	T ₃
The teacher peruses students' records	T ₈	T ₃ T ₅ T ₂ T ₄ T ₃ T ₇	T ₁	

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 maximum⁹ 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).

⁹ 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; Raivoce & 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.

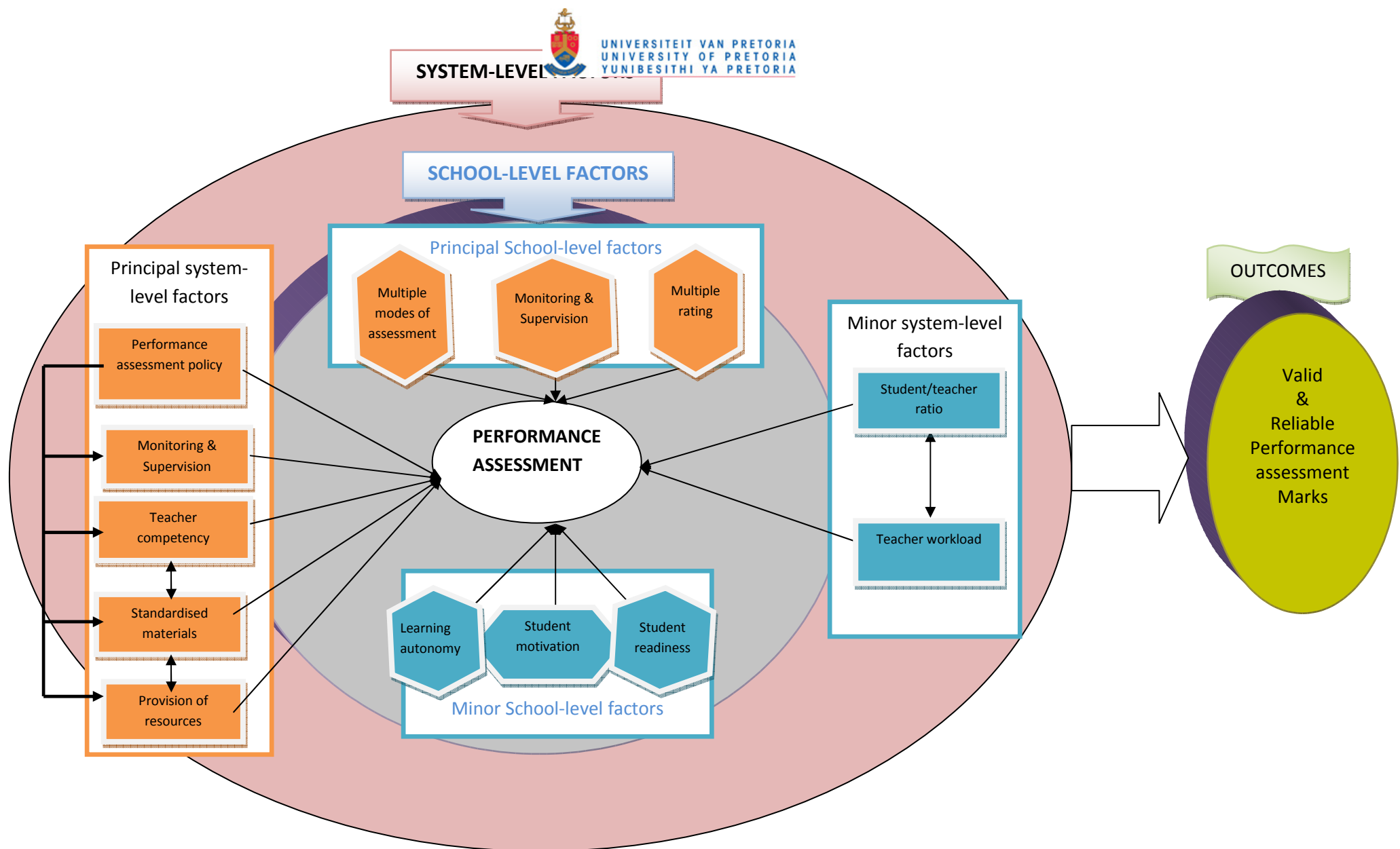


Figure 8.1: Characteristics and quality processes affecting validity and reliability of performance assessment marks

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 (Burris & 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.