

CHAPTER FIVE

AGRICULTURE PERFORMANCE ASSESSMENT PRACTICES IN BOTSWANA

5.1 INTRODUCTION

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

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

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

5.2 BIOGRAPHICAL DATA

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

Table 5.1: The response rate of respondents

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

5.2.1 Teachers' age and gender

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

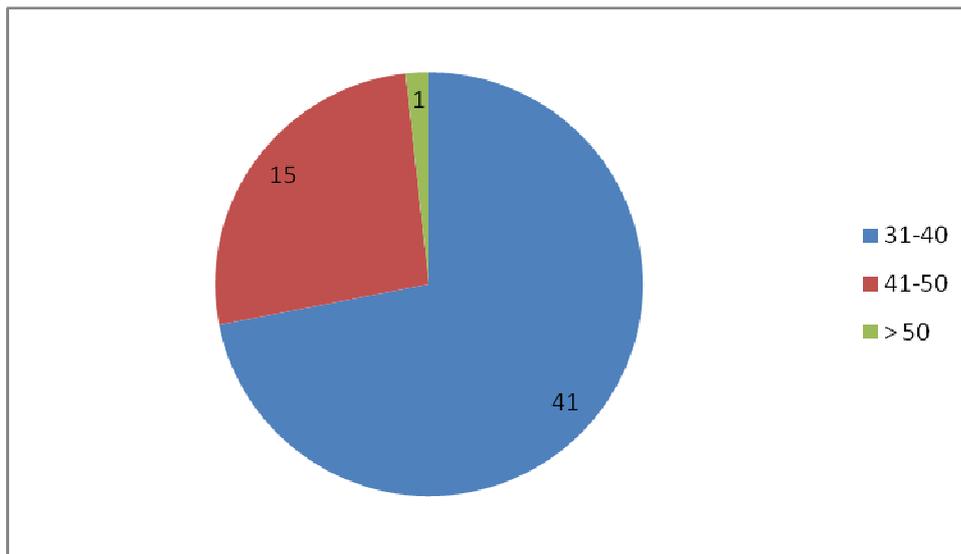


Figure 5.1: Distribution of teachers' age (n = 57)

5.2.2 Teachers' and school administrators' experience

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

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

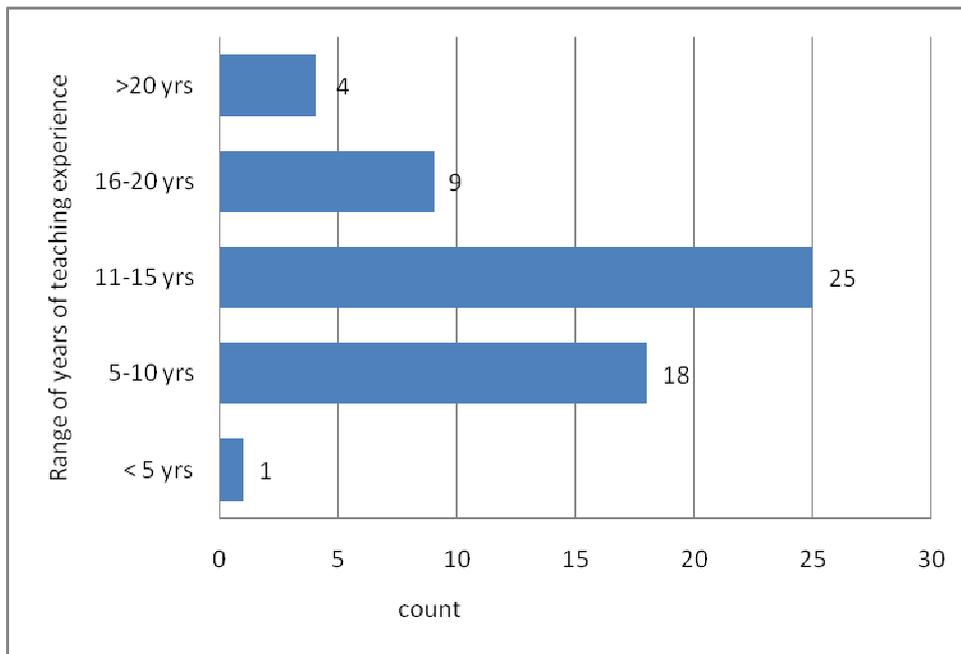


Figure 5.2: Teachers' teaching experience ($n = 57$).

5.2.3 Teachers' and school administrators' qualification and training

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

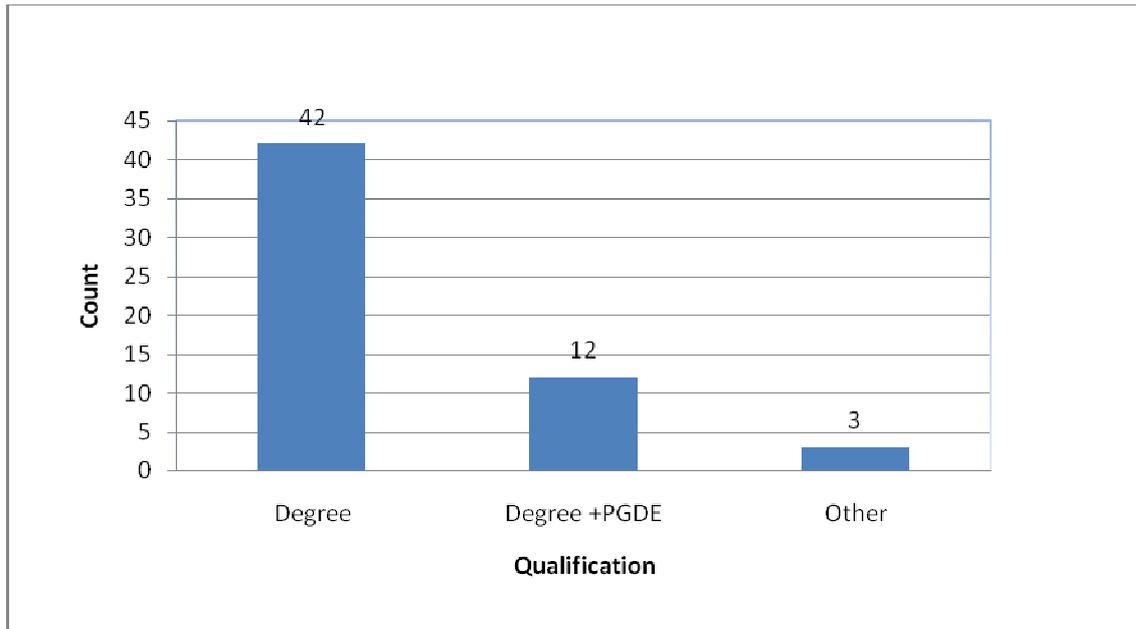


Figure 5.3: Teachers' qualifications (n = 57)

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

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

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

*Applicable to teachers only

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

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

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

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

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

5.2.4 Class size

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

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

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

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

5.3 PERFORMANCE ASSESSMENT PRACTICES OF TEACHERS

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

5.3.1 The mode of assessment

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

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

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

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

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

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

Table 5.6: Characteristics of factors for modes of assessment

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

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

5.3.2 Learning autonomy

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

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

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

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

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

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

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

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

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

5.3.3 Assessment for Learning

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

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

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

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

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

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

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

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

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

5.3.4 Availability of Resources

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

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

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

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

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

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

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

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

5.3.5 *Monitoring and Supervision*

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

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

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

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

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

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

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

5.3.6 *Standardisation of marking*

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

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

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

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

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

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

(*n* = 57)

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

5.3.7 *Attitude towards performance assessment*

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

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

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

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

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

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

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

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

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

5.4 DISCUSSION

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5.5 CONCLUSION

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

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

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

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

5.6 IMPLICATIONS FOR DESIGN OF INTERVENTION

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

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

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

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

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

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

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

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

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

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