CHAPTER FOUR

RESEARCH DESIGN AND METHODS

4.1   INTRODUCTION

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

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

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

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

4.2   PARADIGM UNDERLYING THIS STUDY

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

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

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

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

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

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

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

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

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

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

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

4. 3   OVERVIEW OF RESEARCH DESIGN

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

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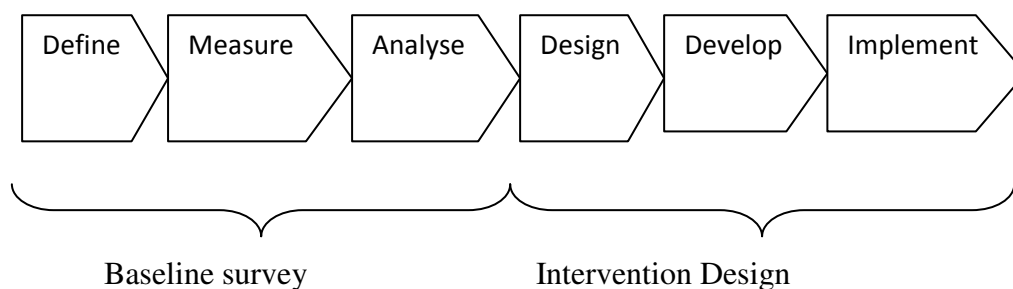

Figure 4.1: The DMADDI approach of DFSS (Source: Islam, 2006. p.52)

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

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

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

The development of prototypes was carried out in collaboration with stakeholders and practitioners (Collins, Joseph & Bielaczyc, 2004; Gravemeijer, 1998; Hoadley, 2002; The Design-Based Collective, 2003), to increase adoption. The development of each prototype is fully discussed in Chapters Six and Seven. Formative evaluation was an integral part of the intervention development and feedback was incorporated into the redesign to improve successive prototypes. Design research was considered appropriate for this study because of the flexibility in developing an intervention stage by stage within the problem of the context, facilitating understanding of implementation problems that practitioners experienced.
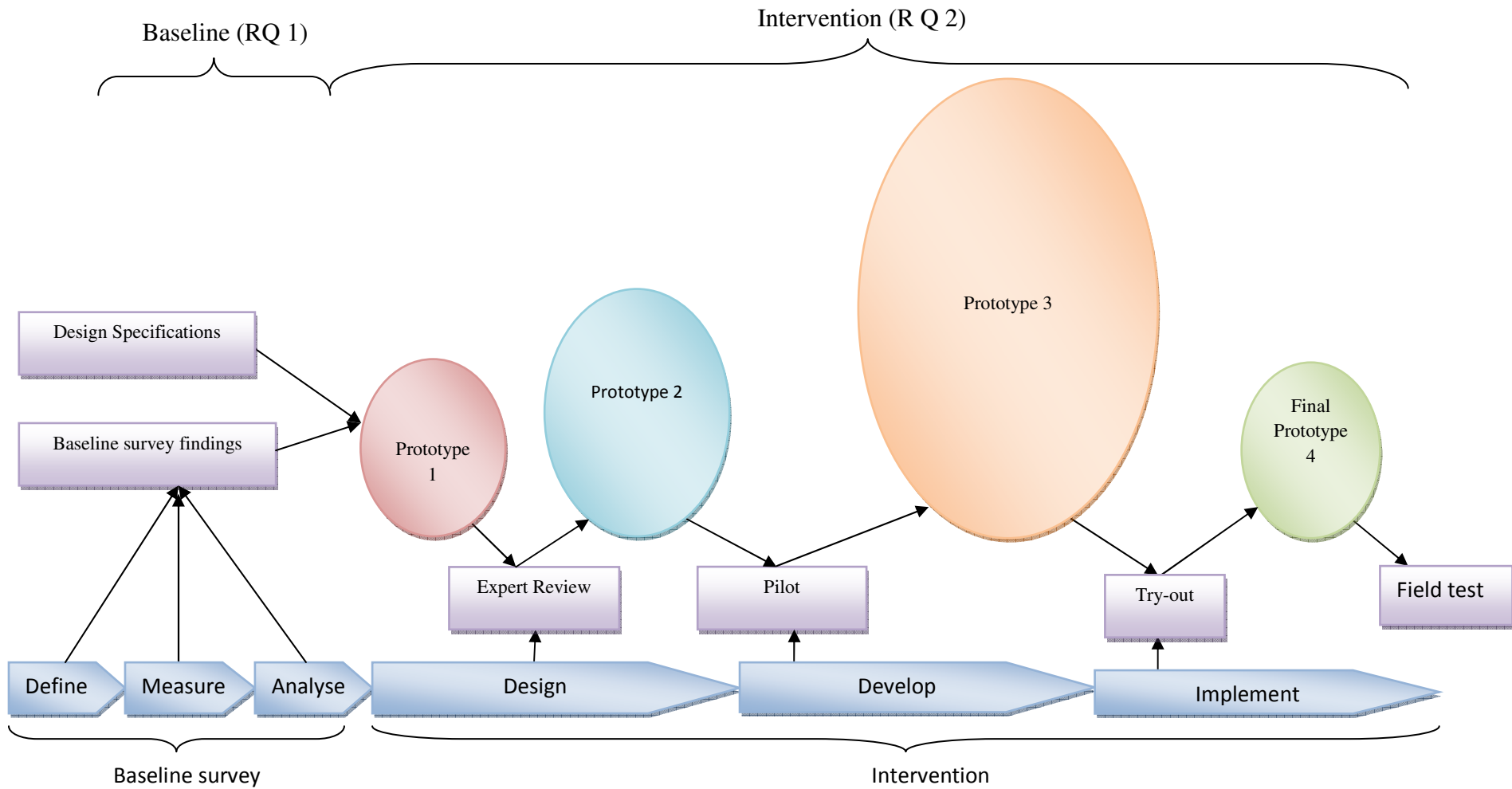
Baseline (RQ 1)

Intervention (R Q 2)

Prototype 3

Design Specifications

Prototype 2

Baseline survey findings

Prototype 1

Final Prototype 4

Expert Review

Pilot

Try-out

Field test

Define

Measure

Analyse

Design

Develop

Implement

Baseline survey

Intervention

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

4.4   RESEARCH DESIGN FOR BASELINE SURVEY: PHASE ONE

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

4.4.1   Research design

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

4.4.2   Research methods

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

4.4.2.1 Sample and Participants

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

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

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

Table 4.1: Sample of participants in the study

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

4.4.2.2 Instrument Development and data collection strategies

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

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

---

[7] For the purposes of this study, school administrators were considered to be school head and deputy school head

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

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

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

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

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

4.4.2.3 Data Collection procedure

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

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

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

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

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

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

Respondents	Sampled respondents	First return rate	Second return rate
School Administrators	26	12	21
Teachers	68	57	57

4.4.3 Data Analysis

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

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

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

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

4.5 RESEARCH DESIGN FOR THE INTERVENTION STUDY: PHASE TWO

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

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

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

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

4.5.1 *The nature of design-based research*

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

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

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

The application of design research in this study drew heavily on the robust paradigms of pragmatism and to a lesser extent from that of constructivism (Walker, 2006), to help understand the relationships between educational theory, designed artefact, and practice (DBRC, 2003). The study remained flexible throughout to accommodate the ever-changing nature of natural settings (Mertens, 2010; Ornstein & Hunkins, 1993). Consequently, the development of the intervention was interactive and responsive to iterative stages of formative evaluation and re-designs (Bannon-Ritland, 2003), which involved multiple design-test-revise cycles as illustrated in McKenny's (2001) CASCADE-SEA study presented in Figure 4.3.
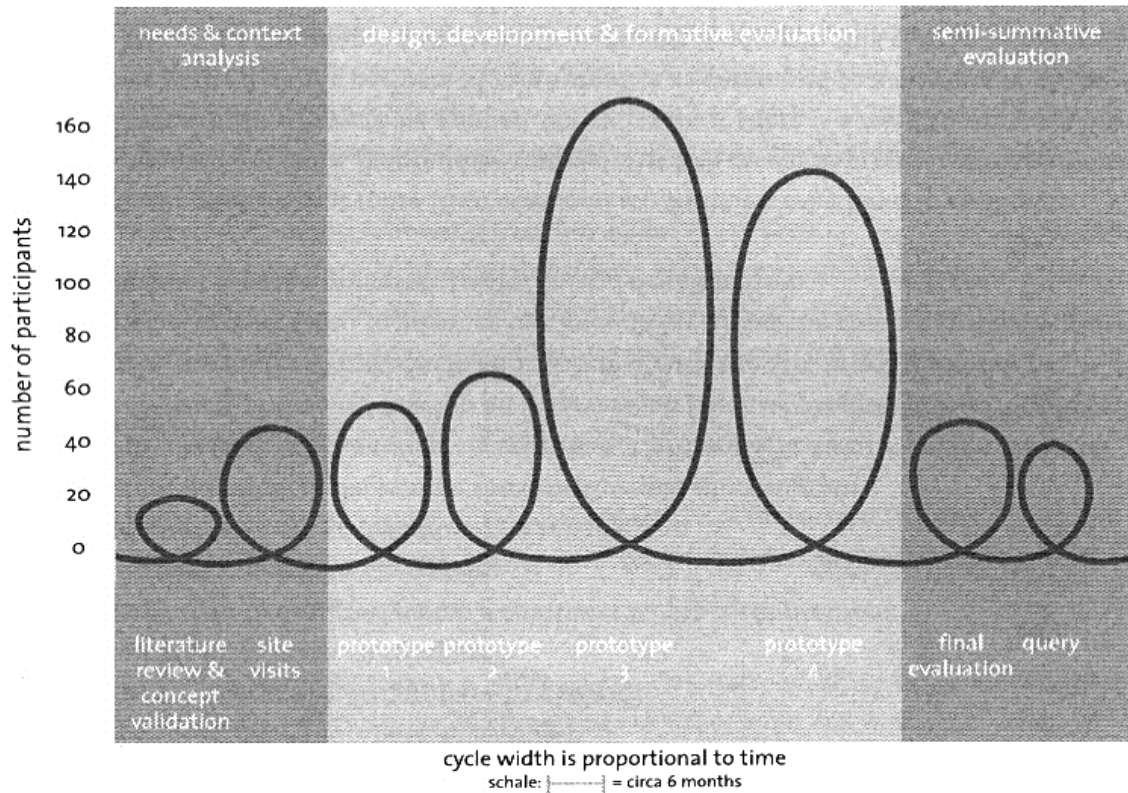
Figure 4.3: The cyclic process of design based research of the CASCADE-SEA study (source: Plomp & Nieveen, 2007. p.14 adapted from McKenny, 2001)

Formative evaluation data helped to refine the initial prototypes and, in turn, to develop a more detailed design intervention (Collins et al., 2004) which applied to other settings for generating design knowledge or principles grounded in broader contexts. Formative evaluation thus has served different functions in the various stages of development and was built within each criteria of quality intervention. Formative evaluation can be perceived as having various layers in a design research study, as illustrated in Figure 4.4 (below). The evaluation comprised four layers, increasing in complexity from bottom to top.
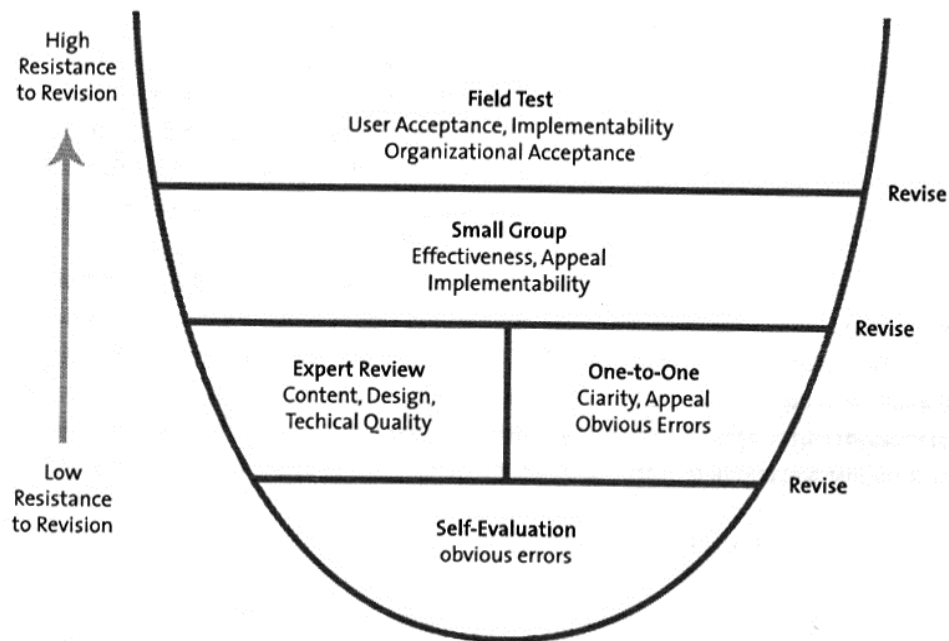
Figure 4.4: Tessmer's layers of formative evaluation (Source: Plomp & Nieveen, 2007.p.28 adapted from Tessmer)

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

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

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

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

Table 4.3: Criteria for high quality intervention

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

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

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

4.5.2   Research design

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

4.5.3   The research process

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

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

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

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

Table 4.4: Schedule of task implementation in schools

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

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

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

4.5.4   *Data collection*

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

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

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

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

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

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

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

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

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

---

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

4.5.5 *Data Analysis*

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

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

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

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

4.6   METHODOLOGICAL NORMS

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

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

4.6.1   *Dependability of the Results*

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

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

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

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

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

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

4.6.2 *Ethical considerations*

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

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

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

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

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

4.7   CONCLUSION

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

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