

## **CHAPTER 3**

# **LITERATURE REVIEW AND CONCEPTUALISATION OF THE STUDY**

## CHAPTER 3

### LITERATURE REVIEW AND CONCEPTUALISATION OF THE STUDY

*This chapter presents the literature relevant to a discussion of assessment strategies in Physics and the conceptual framework of the study. It starts by presenting the research questions and the sources used to find out what is already known about the research questions (3.1). The method used to review the literature, as well as the keywords used to search the information, are also presented in this section. The definition of terms used in this research, and a number of different learning theories used as a platform to guide the discussion are presented and discussed in Sections 3.2 and 3.3 respectively. Section 3.4 begins with presenting arguments in favour of constructivism as central to Science learning (3.4.1). It also presents and discusses arguments about the roles of both teachers (3.4.2) and students (3.4.3) in assessment, and contexts within which different assessment strategies can be considered (3.4.4). The contributions from various authors who worked within the African context are presented and discussed in Section 3.5 as methodological illustrations for the educational design research approach to research on interventions. Drawing from the literature review findings and contextualising the problem posed by the research questions, Section 3.6 presents the issues used to address the main research question adequately.*

#### 3.1 Introduction

The research questions of this study are: What assessment strategies do secondary school Physics teachers in Mozambique apply, and how can they be improved? In order to find an answer to these questions, two research phases were considered. Firstly, the study investigated through a survey approach the assessment practices used by secondary school

teachers in schools. The research question addressed by the Baseline Survey, as referred to in Chapter 1, is:

**What assessment strategies do Grade 12 teachers in Physics in Mozambique apply and what can be said about their quality?**

The second component of the study was characterised by an intervention approach aimed at improving some of the assessment practices used by Physics teachers in schools. The Intervention Study addressed the following research question (refer also to Chapter 1):

**How can teacher assessment strategies be improved?**

Both the Baseline Survey and the Intervention Study were informed by the literature review.

To find out what is already known from the literature about the two study phases, various sources were reviewed both in printed form and electronically. General open source software articles and information were used in order to gain a global idea of the current assessment strategies used by Grade 12 teachers in Physics, as well as the quality of these strategies. Keywords like ‘assessment strategies’, ‘assessment toolkits’, ‘science toolkits’, and ‘mechanical Physics and assessment’ (for Grade 12 or its equivalent) were used to obtain information on what is known about the research question addressed by the survey approach and what the relevant arguments of several scholars are on assessment for learning Physics in general. Thus, the information obtained from the review of these keywords was used to inform the Baseline Survey, particularly in terms of the characteristics of assessment strategies and teaching practices.

The research question addressed by the intervention component of the study was also informed by the keywords referred to above, but the following additional search terms were also used: ‘performance assessments’, ‘formative assessment’, ‘summative assessment’, ‘prototypes’ and ‘demonstration experiments in secondary education’. The term ‘demonstration experiments’, is included in this list because, in Science education

literature, it has been used with different interpretations. Some authors have been using it with the same meaning as ‘laboratory experiments’. This term in this dissertation, refers to activities in the Physics classroom in which students observe, carry out small experiments, interpret phenomena or events occurring to objects, and report the findings, and are all guided by a teacher and teaching materials. In general, relevant considerations following from these keywords are that the future of assessment nowadays depends on the ability to assess student skills in performing tasks; on the power of formative assessment in monitoring student learning; on the role of summative assessment for accountability purposes; and on the importance of demonstration experiments in science subjects (Airasian, 2001; Black & Atkin, 1996; Gardner, 2006; Popham, 2002; Race et al., 2005). Besides using scientific literature from the academic libraries of the University of Pretoria and the Eduardo Mondlane University, other sources were explored by using the following search engines on the internet:

- Academic Information System (University library);
- CAB abstracts;
- ERIC;
- Google scholar;
- ISI web of science;
- Science direct; and
- Tucks (electronic journals for which the University of Pretoria has a subscription).

Table 3.1 indicates which keywords were used in which database search engines.

**Table 3.1: Database search engines**

	Academic Information System	CAB abstracts	ERIC	Google scholar	ISI web of science	Science direct	Tucks
Assessment strategies	✓	✓	✓	✓			✓
Assessment toolkits	✓	✓	✓	✓			✓
Formative assessment	✓	✓	✓	✓			✓
Demonstration experiments in secondary education		✓		✓	✓	✓	
Mechanical Physics and assessment				✓	✓		
Prototypes and Physics			✓	✓	✓	✓	✓
Performance assessments	✓			✓			✓
Science toolkits			✓	✓	✓	✓	
Summative assessment	✓	✓		✓			✓

A significant number of books and articles were found from all these sources. A subsequent selection was made on the basis of years of publication, the education level, and the context. Although the review considered non-African contexts (especially European and American), the focus was on those references reporting research undertaken within and about an African context, or with some similar characteristics, written during the period 1996-2006, and on Grade 12 in Mozambique.

During the review of the databases and search engines, it emerged that some relevant authors were quoted frequently. As the central areas of this research are classroom assessment (both formative and summative), performance assessment, and Physics learning, the literature review focused on authors in these areas, such as: Airasian (2000, 2001), Black (1998), Black and Atkin (1996), Black et al., (2003), Dekkers (1997), Harlen (2006), Kathy and Burke (2003), McMillan (2001), Moskal (2003), Muller (2006), Mutimucuo (1998), Popham (2002), Race et al., (2005), Stiggins (1987), Treagust et al., (1996), and Weeden et al., (2002).

Current writings (such as theses) in the field of Science teaching and learning and reporting investigations within the African context and elsewhere often follow an educational design research approach. As this approach is relevant when addressing the intervention research question, a number of dissertations (viz. Mafumiko, 2006; Motswiri, 2004; Ottevanger, 2001; Teclé, 2006) have been included in the literature review as well.

Given the research questions for this study, the literature review focused specifically on issues related to the ‘assessment strategies’ used for Science subjects, the type of ‘learning evidence’ in ‘alternative’, ‘authentic’, ‘formative’, and ‘performance’ kinds of assessment, and the role of both teachers and students in assessment. In summary, all of the authors reviewed emphasise the need to improve the current teachers’ assessment strategies in schools.

Having presented the research questions being addressed by both the Baseline Survey and the Intervention Study, as well as the sources for the literature review, the next sections deal with: definitions of terms mostly used in the dissertation (3.2); the arguments of several authors on the theoretical orientation of assessment in education (3.3); the roles of teachers and students in assessment; and the need to improve current assessment strategies in Science, with particular emphasis in Physics teaching and learning (3.4).

### **3.2 Definition of terms**

In this dissertation, several key terms are frequently used. For the purpose of the present study and for the reader’s convenience, definitions of key terms used in this study are presented in this section. However, before doing this, it is relevant to clarify the distinction between assessment *for* learning and assessment *of* learning. According to Black et al., (2003) assessment *for* learning is any assessment where the first priority is to serve the purpose of promoting student learning. This kind of assessment is usually informal, embedded in all aspects of teaching and learning, and conducted differently by different teachers as part of their own individual teaching styles. Assessment *of* learning is for grading and certification, occurs in formal settings or rituals, involves non-frequent

tests, is isolated from normal teaching and learning, is carried out on special occasions, and is conducted by methods over which individual teachers have little or no control. This study is about assessment *for* learning.

In this section, the following concepts and terms will be defined and discussed namely: *assessment, authentic assessment, classroom assessment, formative assessment, paper-and-pencil tests, peer-assessment, performance assessment, portfolios, projects* and *prototypes*.

### **Assessment**

The term ‘assessment’ is defined in the Glossary of the 1999 Standards for Educational and Psychological Testing as “any systematic method of obtaining information (from tests and other sources) to draw inferences about characteristics of people, objects or programs” (Chatterji, 2003). Airasian (2001) defines assessment as the process of collecting, synthesising, and interpreting information to aid in decision-making. For this author, assessment involves more than administering, scoring and grading paper-and-pencil tests, and includes the full range of information teachers gather in their classrooms. This information helps teachers to understand their students, monitor instruction and establish a viable classroom culture. Given the two definitions, the researcher’s perception of assessment is that, when the term is used to refer to assessment processes, the activities include writing items or designing an assessment tool, making observations or gathering data using an assessment tool, scoring responses from an assessment tool, developing a scale with specified properties, and administering an instrument using prescribed guidelines. Briefly, the term assessment in the present study is used to refer to a variety of ways teachers gather, synthesise, and interpret information.

In this study, all kinds of assessments used by teachers in schools, from the traditional approaches of paper-and pencil tests to a more constructivist and dynamic process of gathering information following some prescribed guidelines are termed assessment practices or assessment strategies.

### **Authentic assessment**

This is the form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills. McMillan (2001) for instance, defines authentic assessment as an assessment that is constructed to be more consistent with what people do in situations that occur naturally outside classroom. An authentic assessment usually includes a task for students to perform and a rubric by which their performance on the task will be evaluated. In order to determine whether authentic assessment is successful, the school must ask students to perform meaningful tasks that replicate real world challenges to see if students are capable of doing so (Wiggins, 1993). The Mozambican curriculum goals stated in the Physics Syllabus for Grades 11 and 12 emphasise that “the starting point for students’ knowledge acquisition is practical work and the nature and its phenomena is the stepping stone for proving any formulated hypothesis” (MinEd, 1997:1). This shows that authentic assessment drives the curriculum. Teachers have to determine the tasks that students will perform to demonstrate their mastery, and then a curriculum is developed that will enable students to perform those tasks well. Some educators (Meyer, 1992; Stiggins, 1987) distinguish authentic assessment from performance assessment by defining performance assessment as performance-based, but with no reference to the authentic nature of the task.

### **Classroom assessment**

Popham (2002) defines ‘classroom assessment’ as the assessment that comprises a number of assessment decisions taken by the teacher during the teaching and learning process. These decisions occur mainly within the classroom environment, or are informed by the classroom climate. In some educational systems the term ‘classroom assessment’ is referred to as ‘non-standardised assessment’ meaning that the assessments are constructed by teachers, specifically for classroom use, and focused on the particular type of instruction provided in that classroom (Airasian, 2001). The information from classroom assessment is used to provide feedback about the performance of students in a single class, not of students in other classes. In the present study, the term classroom assessment is used to refer to all kind of assessments undertaken by the teachers in the classroom during



the instruction process. Therefore, the terms ‘classroom assessment’ and ‘teacher assessment’ are used interchangeably.

### **Formative assessment**

According to Black et al., (2003:122) “formative assessment is a process, one in which information about learning is *evoked* and then *used* to improve the teaching and learning activities in which teachers and students are engaged”. This definition takes the idea of formative assessment beyond the ‘micro-summative’ assessments of classroom tests and homework. It broadens the sources of evidence and solidifies the notion of what should be done with the evidence. The sources from which evidence can be drawn do not exclude information gathered from formal assessments, but more important than the source of evidence is the idea that the information obtained affects subsequent teaching and learning activities.

### **Paper-and-pencil tests**

Paper-and-pencil tests are assessments in which students write their responses to questions or problems (Airasian, 2001). Examples of paper-and-pencil tests are essays, multiple-choice tests, written assignments, written reports, a drawn picture, or a filling in of a worksheet. In general, paper-and-pencil tests are of two types: selection type – where the student responds to each question by selecting an answer from the choices provided, and supply type – which requires a student to produce or construct a response to a question or task.

### **Peer-assessment**

Generally, peer-assessment is an assessment of the work of others by people of equal status and power. The term can also be used to describe approaches to accountability carried out on behalf of government agencies. In the context of student learning, peer-assessment may be divided into the giving and receiving of feedback and making formal estimates of worth of other students’ work (Brown et al., 1997). It usually involves an element of mutuality and, beneath the processes of giving and receiving feedback, there are implicit criteria of what counts as ‘good’ for different purposes and contexts. The

students' notions of 'good' and 'poor' can be generated in large group classes and tutorials and so provide a basis for reflective learning and more formal peer-assessment. Thus, peer-assessment is rather a tool for learning than a tool for summative assessment.

### **Performance assessment**

The length of responses in items where the student has to supply a response can vary substantially. When, for instance, a student is required to produce complex constructions such as Science experiments reports, book reviews and/or class projects, these assessments are termed performance assessments. As the term suggests, performance assessment requires a demonstration of student skills or knowledge (Moskal, 2003). Performance assessment can take many different forms, which include written and oral demonstrations and activities that can be completed by either a group or an individual. A factor that distinguishes performance assessments from other extended response activities is that they require students to demonstrate the application of knowledge to a particular context. Through observation or analysis of a student's response, the teacher can determine what the student knows, what the student does not know, and what misconceptions the student has with respect to the purpose of the assessment.

### **Portfolios (in education)**

Portfolios, in general, constitute the chief method by which certain professionals such as models, artists, photographers, architects, and journalists display their skills and accomplishments. In the domain of education, a portfolio is a systematic collection of a selected student work (Popham, 2002). It engages students in assessing their own progress or accomplishments over time and establishes ongoing learning goals. A portfolio is not an unrelated collection of student's work but contains consciously selected examples of work that are intended to show student's growth toward important learning goals.

## **Projects**

Projects are purposeful activities of a student or a group of students with a limited duration. They can be finished products which include all written tasks that the students do and which reflect a certain development process of collecting, interpreting and reporting data (Brown et al., 1997). Thus, they are usually carried out in the final year of the course but also may be used in the first year of the course to encourage students to become active, independent students. Projects may be laboratory-based, library-based, work-based, studio-based or community-based. The main purpose of the projects is to develop enquiry-based student skills. While projects enables a student to explore deeply a field or topic, develop an initiative, provide personal ownership of learning and enhance time- and project-management skills, they have a disadvantage of being time consuming to set up, monitor and provide feedback.

## **Prototypes**

A prototype is a model upon which other similar materials are based. It represents all products that are designed before the final product is constructed and fully implemented in practice (Nieveen, 1999). In its initial stage, a prototype can be developed, discussed and modified as required to build consensus. It is, therefore, designed with particular care. In the process of developing a prototype, developers come to an agreement on what to show and how to show it. In the context of the present study, prototypes are physical exemplary materials on teaching and assessment strategies for teachers to use in the classroom that demonstrate the acquisition of student skill and are based on a set of established performance criteria.

This section has presented and discussed the terms or concepts frequently used in the dissertation. The next Section (3.3) provides a number of perspectives of different authors on assessment, which serve as a theoretical orientation to the study.

### 3.3 Theoretical orientation in assessment

In this section, three aspects are regarded as being relevant for providing a theoretical orientation in classroom-based assessment for this study. The aspects are: (i) the objective of assessment and the process of giving feedback to students; (ii) the need for teachers to conduct effective assessment *for* learning; and (iii) the teachers' preparedness on conducting assessment that can generate evidence of authentic learning.

In terms of the first aspect and according to several authors (Black, 1998; Black et al., 2003, Kathy & Burke, 2003; Lin & Gronlund, 2000), assessment may be conducted to serve different purposes, such as assessment to satisfy demands for public accountability; assessment to report an individual's achievements; and assessment to support learning. The focus of assessment in this study falls within the latter purpose (supporting learning) because the study aims to improve those assessment practices that teachers apply. The rationale of focusing on this purpose is that the main aim of schools is to promote student learning and the teacher needs constant information about what the students know. Ideally, assessment should provide short-term feedback so that obstacles can be identified and tackled at an early stage in the learning process. This is particularly important where the learning plan is such that progress with one week's work depends on a grasp of the ideas discussed in the previous week. This type of assessment aims at improving learning, and is called formative assessment or assessment *for* learning.

It is clear that this assessment is the responsibility of the classroom teacher, but others, inside and outside the school might support this work by providing appropriate training and methods for conducting such an assessment. In Science subjects like Physics, however, evidence that a formative assessment is really improving learning must be accompanied by a type of assessment where students are asked to perform real-world tasks and demonstrate the meaningful application of knowledge and skills. This leads to what McMillan (2001) calls authentic assessment and its success depends very much on support that the teacher must receive from various educational stakeholders inside and outside the school. Therefore, in providing such support to teachers, Nuttall (cited in Kathy & Burke,

2003), argues that it is relevant for teachers to know how to generate evidence of authentic learning. Authentic learning, as the Physics Syllabus for Grades 11 and 12 recommends, is crucial for learning science and specifically experimental subjects like Physics. Nuttall (1987) also describes a number of criteria for tasks that validly assess learning, namely: (i) tasks that are concrete and within the experience of the individual; (ii) tasks that are presented clearly; and (iii) tasks that are perceived as relevant to the current concerns of the student. The value of these tasks, in researcher's opinion, allows students to demonstrate good performance because they promote interaction between students and the teacher. In addition, they allow the teacher to get into the students' thinking and reasoning and to evaluate their potential.

Bell and Cowie (2001) distinguish between two types of formative assessment, namely planned formative assessment and interactive formative assessment. These authors suggest that planned formative assessment is used to elicit permanent evidence of students' thinking, and such assessment occasions are semi-formal and may occur at the beginning and end of a topic. A specific assessment activity is set for the purpose of providing evidence that is used to improve learning. All the information is elicited through the task set and the teacher and the student act on this information with reference to the topic itself, with reference to the students' previous performance, and with reference to how the students and the teacher are proposing to take learning forward. Interactive formative assessment is described by Bell and Cowie (2001) as taking place during student-teacher interaction. This refers to the incidental or ongoing formative assessment that arises out of learning activity and cannot be anticipated.

As is the case with planned assessment, in interactive formative assessment the purpose is to improve learning by mediating the student learning. The process involves the teachers noticing, recognising and responding to students' thinking and it is more teacher- and student-driven than curriculum-driven. Unlike the kind of permanent information that accrues from planned assessment, this kind of assessment generates information that is ephemeral. The latter kind of formative assessment is crucial for this study because it is

important for enhancing student learning, and therefore, the teacher must be supported in knowing how to react in relation to what is deemed at the time to be worth noticing in the student. Unlike in the planned formative assessment where there is a longer time gap in responding, in the interactive formative assessment, the teacher's response is immediate, and the kind of planning that can still be made is on how to facilitate dialogue and tasks between him/her and the students. In an interactive assessment, students are given opportunity to argue about the assessment tasks and to challenge teachers' responses to their questions.

As for the importance of immediate and ongoing feedback, Race et al., (2005) elaborate on how quality feedback can best be given to students. Amongst the several aspects of quality feedback referred to by these authors, they mention the following aspects of quality feedback: (i) time - the sooner the feedback is given the better; (ii) personality - it needs to fit each students' achievement; (iii) expressed - whether congratulatory or critical; and (iv) empowerment - both congratulatory and critical feedback must not dampen learning, but rather strengthen and consolidate it.

In conclusion, for the assessment objectives and feedback, three major aspects provide orientation to the review on assessment for this study. Firstly, the assessment is carried out to support learning, therefore, the provision of feedback should be on a short-time basis so that obstacles in the learning process can be tackled in good time (Black, 1998; Race et al., 2005). Secondly, teachers need to have at their disposal certain students' tasks that can validly assess particular learning and generate evidence of authentic learning (Kathy & Burke, 2003; Lin & Gronlund, 2000; Popham, 2002). Thirdly, immediate and ongoing feedback is crucial to facilitate student-teacher interaction (Bell & Cowie, 2001; Race et al., 2005).

With reference to carrying out an effective assessment *for* learning, it is worth mentioning that it requires having students actively engaged in finding solutions to problems they face and developing the ability to construct knowledge. In this process, the role of the teachers as facilitators is crucial in monitoring the assessment practice. James (2006), on the relationship between assessment practice and the ways in which the processes and

outcomes of learning are understood, argues that three theories of learning and their implications for assessment practice can be distinguished. These are discussed below.

- **Behaviourism:** this is where the environment for learning is the determining factor, the learning is the conditioned response to external stimuli, and rewards and punishments are the powerful ways of forming or eradicating habits. The implications for assessment practice are that the progress is measured by timed tests, performance is interpreted as either correct or incorrect, and poor performance is remedied by more practice in the incorrect items.
- **Constructivism:** this is where the learning environment is determined by prior knowledge - what goes on in people's minds - emphasis is on 'understanding', and problem solving is the context for knowledge construction through deductive and inductive reasoning. The implications for assessment are that self-monitoring and self-regulation are relevant dimensions of learning, and the role of the teacher is to help 'novices' to acquire 'expert' understanding of conceptual structures and processing strategies to solve problems. When students are involved in the construction of their own learning through formative assessment, they develop the ability to monitor and regulate their learning agenda.
- **Socio-culturalism:** this is where learning occurs in an interaction between the individual and the social environment. Thinking is conducted through actions that alter the situation and the situation changes the thinking. The implication is that, prior to learning, there is a need to develop social relationships through language, because it represents the central element to our capacity of thinking.

It has been argued that the latter theory is not yet well worked out in terms of its implications for teaching and assessment (James, 2006). Teaching and learning tasks need to be more collaborative and students need to be involved in the generation of problems and of solutions, because the current perspective of assessment within this perspective is still inadequately conceptualised. For the context of this study, the constructivist theory of learning is recommendable. The reason is that the Mozambican teaching system emphasises the importance of considering children's prior knowledge before helping them understand other conceptual structures. The implication of this choice for assessment is

that this construction of children's own learning can be easily facilitated through formative assessment advocated by this study.

Regarding teachers' readiness to conduct assessment, i.e., what teachers need to know in order to assess learning and to generate evidence of authentic learning, James and Pedder (2006) explain that, as a pre-condition to enhance assessment for learning, changing pedagogical practice should be taken into account, particularly with the roles of both teachers and students. Teachers should be supported in developing skills to plan assessment, interpret learning evidence, and provide feedback to students and support them in assessing their work and that of their peers. This means that teachers need to practise new roles and try and evaluate new ways of thinking. Students should also be helped to take on new roles as students. They should understand the learning goals and identify the criteria used for assessing their progress, develop skills of peer and self-assessment, and make progress through constructive formative feedback from peers and their teacher. This implies developing a language and eagerness for talking about teaching and learning.

Within this framework, the present literature review seeks, amongst other issues, to understand how teachers, who have been trained following a behaviourist theory of learning (Buendía Gomez, 1999; Siteo, 2006), can facilitate student learning (and assessment) in a constructivist approach, as advocated by the Mozambican syllabus and recommended by recent literature, without neglecting socio-cultural relationships. In fact, for improved Physics learning, an effective formative assessment carried out in conjunction with authentic assessment, as argued earlier on, can better be achieved if implemented taking into account different learning theories. Students learn Physics better when their prior knowledge is taken into account and when their ability to perform real-tasks is encouraged (Dekkers, 1997; Mutimucuo, 1998); and this is also the reason why authentic performance assessment is crucial. The combination of formative and authentic assessments is very important because, while the performance of real-world tasks is supported by authentic assessment, the learning of necessary basic concepts and principles is merely dealt with by formative assessment.



### **3.4 Assessment strategies in Physics**

The purpose of this section is to discuss arguments of several authors regarding different assessment strategies used to assess Physics learning in varied contexts. It begins with examining the principles of constructivism theory seen as central theory relating to Science education in general, and to Physics learning in particular. The importance of performance and authentic assessments in enhancing Physics learning is also discussed in this part. The second and third parts discuss the role of the teacher and the role of the students in assessment, respectively. The fourth part presents several arguments of the authors reviewed about different assessment strategies and the context in which they are used. Lessons drawn from this section are summarised in the fifth part.

#### ***3.4.1 Constructivist views of learning***

Constructivism clarifies the views about the nature of human knowing, particularly the nature of scientific knowledge, as well as a view about learning processes and validation procedures of the acquired knowledge, and it is seen as a powerful theoretical resource that maximises student learning (Mutimucio, 1998; Treagust et al., 1996). Both psychological and epistemological principles of constructivism emphasise that knowledge cannot be separated from knowing subject. The epistemological principle states that the function of cognition is adaptive and enables the student to construct viable explanations of experiences of the world. The psychological principle states that students do not passively receive knowledge but they actively build up their knowledge by a cognizing subject (Treagust et al., 1996). When the assessment of the knowledge is taken into account this process of building up knowledge becomes somehow problematic due to the different levels of activities (both manipulative and mental activities) that the students may be involved and to the different realities (realities accessible via sensory organs, via theoretical understanding and via instrumentation) where they are living in. Therefore, in assessment there is a need for consensus about levels of activities and different realities and for this process of knowledge construction, what the student already knows is of central importance. Consequently, knowledge about the world is seen as a human

construction, and this view of a student as an active agent, constructing his/her own reality, determines life processes and changes of all living beings (Mutimucuo, 1998; Piaget, 1972).

Literature on Science education and the perception of the issues pertaining to assessment strategies in Physics were considered to expand the view on student difficulties when it comes to assess the learning. Because constructivism is seen as the theoretical framework on which most research into student thinking and learning is based, the discussion on assessment strategies on Physics, either in Africa or internationally, is based on this theory (Mutimucuo, 1998; Treagust et al., 1996). International literature (Airasian, 2001; Moskal, 2003; Stiggins, 1987) emphasises the role of performance assessment as crucial in assessing Physics learning. But for students to be able to learn Physics better they should not only be asked to do performance tasks. They should primarily be able to understand basic concepts related to the subject. This type of knowledge can be and is mainly assessed through paper-and-pencil tests, but sometimes with other kinds of formative assessment. In fact, available research findings indicate that assessing students' basic skills seems not to be the problem with Mozambican secondary school teachers (INDE, 2005; Lauchande, 2001). According to this literature, paper-and-pencil tests and oral questioning have been used by these teachers with successful results at this respect. The problem is that students are not assessed on their ability to perform real-world tasks, i.e., their skills or proficiency in doing something. This is the rationale of choosing performance assessment for this study as one of the assessment strategies that can help teachers to improve student learning of Physics. This choice will be argued in Chapter 5 as a conclusion from the Baseline Survey. Performance assessments call upon the students to demonstrate specific skills and competencies, i.e., to apply the concepts and the knowledge they have acquired. The form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential skills and knowledge is authentic performance assessment and it is closely related to the constructivist theory of learning. According to Muller (2006) one of the most critical features of an authentic assessment is that it usually includes a task for students to perform with a rubric against which their performance on the task will be evaluated.

On the other hand, Muller (2006) discusses the process of evaluating student performance through rubrics. The author argues that performance assessments are typically criterion-referenced measures where a student performance on a task is determined by matching the student performance against a set of criteria to determine the degree to which the student performance meets the criteria for the task. To measure student performance against a pre-determined set of criteria, a rubric, or scoring scale, is typically created which contains the essential criteria for the task and appropriate levels of performance for each criterion. A rubric is comprised of two components: criteria and levels of performance. For instance, in a task where a student is asked to assemble an electric circuit (manipulative activity), one of the criteria could be the student ability to obtain all the necessary equipment for the task and the other criterion could be the student ability to assemble the circuit properly. For both criteria, it can be defined whether the level of performance is poor or excellent. The rubric describes the task itself, i.e., the assembling of the electric circuit. Each rubric has at least two criteria of good performance and at least two levels of performance at which that performance was achieved. The criteria are the characteristics of good performance on a task. For each criterion, the teacher or the evaluator applying the rubric can determine to what degree the student has met the criterion, i.e., the level of performance. Table 3.2 shows the components of a rubric as well as the criteria and the levels of performance that can be used to assess it.

**Table 3.2: Components of a rubric**

<b>Rubric: Assembling an electric circuit</b>		
<b>Criteria</b>	<b>Levels of performance</b>	
<b>Criterion 1:</b> Student ability to obtain all the necessary equipment.	<b>Poor</b> (e.g., from 0 to 1): Basic equipment necessary for the electric circuit not found	<b>Excellent</b> (e.g., from 2 to 4): All equipment for the assembling of the electric circuit in place.
<b>Criterion 2:</b> Student ability to assemble the circuit properly.	<b>Poor</b> (e.g., from 0 to 1): Electric circuit not accurately assembled and some essential parts missing, leading to a malfunctioning of the circuit.	<b>Excellent</b> (e.g., from 2 to 4): Electric circuit accurately assembled and working properly.

In between the levels of performance ‘poor’ and ‘excellent’ some other levels can be considered namely ‘satisfactory’ and ‘good’ with redistribution of points to fit in the scale of 0 to 4. Each criterion has its own scoring scale (e.g., from 0 to 4 points) and the scale is applied along the different levels of performance with a previously established maximum score (e.g., 20 points). Hence, a rubric appears to be a scoring scale used to assess student performance in terms of a task-specific set of criteria.

In conclusion, both the psychological and the epistemological principles of constructivism emphasise that students need to be actively involved in building up their knowledge. In this process of knowledge construction, the discussion on assessment strategies taking place in Africa and internationally, emphasises the role of performance assessment as being crucial for assessing Science subjects. As was referred to earlier on in this subsection, performance assessment alone cannot help students learn Physics. Other types of assessment strategies - for instance, paper-and-pencil tests, verbal tests, and portfolios - are also important and are needed to assist students to learn the necessary knowledge and skills in order to do a performance assessment. This study, however, focuses on the use of performance assessment because the Baseline Survey findings (see Chapter 5) have indicated that teachers are already achieving some success by using those other types of assessment leading to the successful completion of a performance assessment. Assessment with these characteristics will be argued in Chapter 5 as a conclusion from the Baseline Survey. Performance assessment deals with the role of students on demonstrating their skills and competencies, as well as the role of the teachers in monitoring the learning process and in evaluating the levels of such performances.

The following subsections addresses firstly, the role of the teacher in assessment and secondly, that of the students. Thirdly, arguments of various authors about different assessment strategies applied to different contexts are presented and discussed as platform for designing the improved assessment strategy advocated by this study.

### 3.4.2 *The role of the teacher in assessment*

Before elaborating any further, it is important to explain the need to improve the current assessment practice in schools. Effectively, assessment remains the weakest aspect of teaching and learning in most subjects, including Physics. Nearly all school assessment policies have weaknesses, which are reflected in corresponding weaknesses in the assessment practice of teachers. For example, Weeden et al., (2002) argue that, in many classrooms, the issue is not that teachers are not assessing enough, but that they are not using the information they collect to help students learn.

The problem being pointed out by the present study is linked to the collection of learning evidence and it is twofold: (i) that Grade 12 Physics teachers have been using limited types of assessment (mostly paper-and-pencil tests); and (ii) that also with these limited types, the collected evidence has only been used for accountability, and not for promoting learning. A strong argument of this study is that any support for the role that a teacher can play in the classroom should be directed towards trying to find a solution to this problem.

Race et al., (2005) on the problem of the limited types of assessment used by teachers, for instance, discuss the importance of putting assessment into context and they stress the need for teachers to consider several aspects of assessment. These aspects include knowing why to assess, what to assess, and what the quality is of feedback they should provide to their students. Regarding the issue of *why to assess*, these authors argue that teachers might be supported to understand, amongst other aspects, that assessment is carried out to guide student improvement and help them to learn from their mistakes (formative assessment). In relation to *what to assess*, teachers' role might be to assess the 'process' of how the students are achieving the learning outcomes and in a holistic approach rather than the 'product' (outcome) itself. As for the *feedback quality*, the authors emphasise the need for a timely, personal, expressed and empowering feedback for student learning. The authors' argument is that the use of a variety of assessment types will support and inform student learning. Contextualised learning is in line with a

constructivist approach, but one cannot expect to assess different learning skills with limited assessment types and taking into account different student backgrounds.

In relation to the collection of evidence, Harlen (2006) elaborates on how teachers must be encouraged to collect student learning evidence as a normal part of class work either in an informal or formal formative assessment. In particular, the author highlights the importance of criterion-referenced assessment through which the student's achievement is described in terms of what s/he can effectively do, as opposed to norm-referenced assessment that is based on the ranking of students in order of their achievement. Student- and criterion-referenced assessment must be the basis for judging the evidence, the feedback must be judged and used by both students and teachers, and the assessment in general should be directed *for* learning. The fact is that, according to the available Mozambican literature (INDE, 2005; Lauchande, 2001), current teacher assessment practice in Science subjects has a more formal summative character whereby the collection of evidence is done as a separate task or test. The basis of judgment is criterion-referenced and the assessment is generally an assessment *of* learning.

### ***3.4.3 The role of students in assessment***

The arguments of several authors in Science education (Dekkers, 1997; Harlen, 2006; Race et al., 2005) suggest that an effective assessment for learning strategies depends, among other aspects, on active student involvement, their ability to assess their colleagues and themselves, and on the profound influence assessment has on the motivation and esteem of the students.

Dekkers (1997), for instance, argues that research on student knowledge of the world requires a basis of scientific knowledge that the teacher and students share, as well as effective communication between them. The author is of the opinion that, for the establishment of such a base of knowledge, it is important to start with establishing which knowledge is shared. If there is no such basis, no mutual understanding can develop.

On the issue of student involvement, Race et al. (2005) argue that the provision of feedback either to individual or to groups of students helps to improve their active participation in the learning process. The authors also emphasise that, if a teacher increases the students' participation, he/she must allow them to interrogate and challenge his/her comments. In subjects like Physics, this is crucial because interrogation and the expression of the student thoughts help to mediate their reasoning. Peer-assessment is another factor regarded as relevant by these authors for successful assessment. They argue that students learn more intensely when they have a sense of ownership of the agenda, and by assessing their peers, they learn from each other's successes and weaknesses. Providing or negotiating assessment criteria, gradually introducing peer-assessment, and making peer-assessment marks meaningful (i.e., making the marks count) are, according to the same authors, some of the aspects regarded as being useful for a successful peer-assessment. In fact, meaningfulness of assessment marks is also crucial in increasing students' motivation. On this respect, Harlen (2006) claims that most of the roles that students can play in assessment in particular, and in learning in general, have much to do with their motivation. It represents the construct that impels students to spend the time and effort needed for solving problems and for learning. The author also argues that students do not only gain motivation as an input from education, but it is also an outcome if they are able to adapt to the world of changing conditions that occur beyond formal schooling. When such changes occur rapidly, the motivation of students to learn new skills will be stronger and their enjoyment of encountering new challenges will be greater. Consequently, assessment is seen as one of the key factors that affect student motivation. Still according to Harlen, some authors such as Stiggins (2001), claim that teachers can enhance or inhibit student desire to learn more quickly through their use of assessment than through any other instructional means they can use.

Students' self-esteem is another important factor in learning and assessment (Race et al., 2005). It is defined as the way people value themselves both as people and as students, and shows the confidence that the person feels in being able to learn. This means that any role to be played by students in assessment strongly depends on the level of their own self-

esteem. Those students who are confident about their ability to learn will approach any assessment task with an expectation of success and a determination to overcome problems.

#### ***3.4.4 Assessment strategies and the context***

Airasian (2001) has pointed out that teachers normally use three main strategies to gather their assessment information namely, observation, oral questioning, and paper-and-pencil tests. According to this author, the paper-and-pencil methods is the most important method teachers use to collect assessment information and they are of two general types of methods, namely, selection and supply. In the selection type, students respond to each question by selecting an answer from the choices provided. In supply or constructed response type, the student produces a response to a question or task. In the selection type, the advantage is that it provides the maximum degree of control for the question writer, in a supply-type item, the question writer only has control over the question itself, since the responsibility for constructing the answer resides with the supplier.

The observation method is one of the major strategy teachers use to collect assessment data about students, instruction, and learning. It involves watching or listening to students carry out some activity, or judging a product a student has produced. For example, when students submit a Science project or set up laboratory experiment, the teachers also observe and judge what the students have produced. Both planned and unplanned observations have the advantage of allowing teachers to observe a particular of student behaviour which is thus considered as important information gathering techniques in classrooms.

The oral questioning is another method mostly used by teachers not only to collect assessment information but also for guiding instruction. It can be used to review a previously taught topic, brainstorm a new topic, find out how well the lesson is being understood by the students, and to gain the attention of the disturbed students. The advantage of this strategy is that it allows the teacher to gather information related to



assessment without the intrusiveness of administering paper-and-paper assessments. Formal oral examinations, for instance, are used in subject areas such as foreign language, singing, and speech. Oral questioning techniques are also seen as vital to complement all other information gathering strategies.

Other authors discuss many other assessment strategies. With the portfolios strategy, for instance, it is worth considering arguments by Kemp and Toperoff (1998). These authors define portfolios as collections of student work representing a selection of performance. Portfolios may be a folder containing a student's best pieces, and the student's evaluation of the strengths and weaknesses of the pieces. It may also contain one or more works-in-progress that illustrate the creation of a product, such as an essay, evolving through various stages of conception, drafting, and revision. According to Kemp and Toperoff (1998), recent changes in education policy in the United States, which emphasise greater teacher involvement in designing curriculum and assessing students, have been an impetus to increased portfolio use in schools. They are valued as an assessment tool because, as representations of classroom-based performance, they can be fully integrated into the curriculum. And unlike separate tests, they supplement, rather than take time away from instruction. Moreover, many teachers, educators and researchers believe that portfolio assessments are more effective than 'old-style' tests for measuring academic skills and informing instructional decisions. Popham (2002) distinguishes some advantages and disadvantages of portfolios. Portfolio assessments are difficult to evaluate because they are tailored to individual student's needs, interests, and abilities and they take time to carry out properly. On the other hand, however, they are a way of documenting and evaluating growth and allow student self-evaluation and personal ownership.

Popham (2002) also supports performance assessment. He points out that many teachers consider short-answer and essay tests a form of performance assessment, which means that they equate this kind of assessment strategy with any form of constructed-response assessment. Other authors (Airasian, 2000; Moskal, 2003) contend that genuine performance assessments must have at least three characteristics. These are: multiple evaluative criteria, in which the student performance is judged using more than one

evaluative criterion; prespecified quality standards, where each of the evaluative criteria on which a student performance is to be judged, are explicated before judging the quality of the performance; and judgmental appraisal, where genuine performance assessments depend on human judgements to determine how acceptable a student performance really is. For whatever reasons, many advocates of performance assessment prefer that the student tasks should represent real-world rather than school-world situations (Airasian, 2001; Moskal, 2003; Stiggins, 1987; Wiggins, 1993). Authentic assessment and alternative assessment are phrases used by some authors (McMillan, 2001; Meyer, 1992; Stiggins, 1987) to describe performance assessment. Authentic assessment is so considered because the assessment tasks are more closely linked to real-life and not to school life, while in alternative assessments the tasks are alternative to those of traditional paper-and-pencil tests.

### ***3.4.5 When reading from this section***

From the several arguments presented and discussed in the above section two major lessons emerge as relevant for addressing the main research question of the present study:

Firstly, that a number of assessment strategies can be used to assess Physics learning and the constructivist theory of learning appears to support some of them. Depending on the assessment objectives being pursued (supply answers or constructed elaborated responses, learning by doing, guiding instruction, evaluating student work) and on the context in which the assessment takes place (normal classroom situation, laboratory setting, out-of-school environment) one can make adequate decisions on which learning theory better suit which assessment strategy. For this particular study, constructivism is appropriate to guide students to learn in a laboratory setting with the aim of achieving the goal of learning by doing. In fact, the Baseline Survey of this study was carried out taking into account both the assessment objectives sought by the teachers and the context in which these teachers were assessing their students. The constructivism theory, seen as central for Physics learning, was influential in this study during the design and development of the exemplary assessment materials (see Chapter 4).

A second lesson from the literature reviewed is that the roles of both teachers and students are crucial for the success of any assessment strategy. It is important for the teachers to know why to assess, what to assess, and what the quality is of feedback they should provide to their students. A criterion-referenced student assessment must be the basis for judging the learning evidence, the feedback to students must be judged and used by both students and teachers, and the assessment in general should be directed *for* learning.

For the students' perspective, the literature suggests that an effective assessment depends mainly on their active involvement, their ability to assess their colleagues and themselves, and on their motivation and self esteem. Therefore, any improvement of assessment practices can only be effective if it includes those assessment practices or strategies that emphasise formative approaches as a way of improving student learning. This study then, addressed these aspects during the Intervention Study through an instructional strategy of students' knowledge construction named Predict-Observe-Explain (see Chapter 4, Section 4.3, under design guidelines for the intervention).

Having presented the arguments of various authors on the topic under investigation, the following section reviews the literature on intervention studies conducted within an African context in the field of Science education.

### **3.5 Some intervention studies in science education**

The content of this section is presented through two perspectives, namely the findings of the reviewed intervention studies, and the methodological and substantive implications of such an approach for the present study.

During the review of the literature, some writings related to research on interventions were sourced. Most of this literature was in the form of PhD theses written within an African context and emphasising an educational design research approach as the most suitable for intervention studies. The writings are rooted in the field of assessment in Science

education, and are particularly related to probing students' understanding of Science using this approach. From all the authors reviewed in this field, the findings from research by Mafumiko (2006), Motswiri (2004), Ottevanger (2001) and Tecele (2006) are relevant to the research reported in this thesis.

For example, the aim of Mafumiko's study - *Micro-scale experimentation as a catalyst for improving the chemistry curriculum in Tanzania* - was to investigate the possible use of a low-cost approach to practical work that could contribute to improving the teaching and learning of chemistry in Tanzanian secondary schools. The study focused on designing and evaluating an intervention of micro-scale experiments to support curriculum materials at this level. The main research question addressed by this study was 'what are the characteristics of micro-scale chemistry materials that contribute to the initial implementation of practical work in chemistry education in Tanzanian secondary schools?' The key findings were that (i) teachers and students were able to implement most of the lesson activities according to advice provided in the curriculum materials (classroom implementation); (ii) teachers regarded having access to the support materials in advance as very helpful for preparation of the lessons in general (opinions about the study approach); (iii) teachers considered micro-scale experiments as a useful way of conducting practical work because it enabled them to involve a large number of students with minimum resources (opinions about conducting practical work); (iv) students involved in the micro-scale experiments found them helpful in enhancing their learning of chemistry, made the subject enjoyable and, hence, increased students' participation in the lessons (opinions about the approach); (v) students found their involvement in chemistry micro-scale experiments as increasing their confidence in doing more experiments as well as their awareness on safety and environment (opinions about learning of chemistry). In general, these findings indicated that for the teachers, the materials provided adequate support information during the preparation of students' practical work with the development approach which needed less time, and less sophisticated equipment.

Tecele (2006) in *The potential of a professional development scenario for supporting biology teachers in Eritrea* addressed the question of 'what are the characteristics of a

professional development scenario that effectively supports biology teachers in Eritrea implementing a more student-centred approach?’ Similar to Mafumiko’s study, this study adopted an educational design research approach to guide the analysis, design, evaluation, and revision processes of the professional development scenario (intervention). This study found that teachers regarded prototyping of professional development scenario as being important and useful in providing them support on subject matter knowledge, lesson organisation, using concepts maps, and handling group activities. However, in some cases teachers were observed encountering problems with group work activities and throughout the tryouts, the issue of time continued to be problematic. Teachers needed more time, particularly in drawing conclusions from the activities. In general, teachers appreciated the summative evaluation workshop because it provided them with exemplary materials, a forum for active discussion, the opportunity to observe exemplary practice, and a learning environment for practicing and augmenting the skills for teaching practically-oriented biology lessons.

Motswiri (2004) conducted an investigation on *Supporting chemistry teachers in implementing formative assessment of investigative practical work in Botswana* and addressed the research question of how can exemplary curriculum materials support senior secondary chemistry teachers in Botswana with the implementation of formative assessment of students’ investigative practical work. This study also followed an educational research approach where a prototyping process was used for an orientation study aimed at (i) articulating initial design specifications for the envisaged exemplary materials, (ii) developing and trying out several versions of prototypes, and (iii) field testing the final version. Findings from this study indicated that teachers were critical about the congruence of the exemplary materials (the intended practice appeared to be incongruent with the teachers’ current practice), but they were positive about their clarity (the materials were regarded understandable) and their cost (the suggested implementation was possible within the limitations of available resources in the science laboratories). However, the teachers were not often observed to demonstrate formative assessment orientations in terms of asking questions related to helping students to reflect critically on results they expected, activities they carried out, and results they obtained. They seemed to

be in need for more support in terms of formative assessments with particular emphasis on time management. Their use of practical work was more frequent in the body of the lesson (where students were helped with probes to participate in planning and experimenting) than in the lesson introduction and lesson conclusion. Students, in contrast, enjoyed the lessons, especially during lesson introduction and learned subject-specific knowledge in terms of investigation procedures.

Ottevanger (2001) on *Teacher support materials as a catalyst for science curriculum implementation in Namibia* addressed the question of what are the characteristics of materials that adequately support teachers in the initial implementation of Science curriculum innovation in the classroom. With the same research approach used in the other studies, Ottevanger's study found that (i) from the scientific process point of view the teacher support materials have led to well-organised lessons in the majority of cases and were useful as a resource in offering extra information on the topic of the lesson. This was seen by the author as a positive step forward in the Namibian context. (ii) The connection between the specific experiment and the relevant theory needs to be further strengthened. (iii) Students' involvement in the lessons increased during the lessons supported by the developed materials. (iv) Students indicated that they liked using materials from local context, doing group work and cooperating with other students. They also referred to the fact that their teacher appeared to be better prepared than in their usual classes. (v) Although teachers seemed to address the time issue in their own ways, this appears to be a continuous problem in completing lessons. In conclusion, this author claims that in the context they were used, teacher support materials containing procedural specifications have shown themselves able to act as a catalyst in the initial implementation of the new curriculum in the classrooms.

The relevance of all these interventions for the present study can be described from both methodological and substantive perspectives. Although the focus of all these studies was on improving student-centred learning, and not particularly on assessment strategies, their methodology can still be successfully applied to the present study. There are similarities between the present study and the literature discussed that support this argument. Firstly,

all studies focused on design and formative evaluation of exemplary materials where the search for characteristics of an effective intervention was conducted while teachers and students were working on that intervention. Secondly, there was a decision to focus on a certain topic or theme to concentrate on. Thirdly, the tasks carried out either by students or by teachers were designed in a standardised manner. Fourthly, the methodology included anticipation of potential implementation problems through application of a systematic process of formative evaluation of the products. Fifthly, the support materials for teachers were designed in such a way that they provided help at four support levels, namely, subject knowledge, lesson preparation, teaching methodology, and assessment and feedback. All these aspects are concurrent and characterise the methodology of research on interventions.

Substantively, from all the contributions and arguments presented in the reviewed studies, the emphasis is on the importance of developing teacher support materials and on the design and tryout of authentic material in a classroom environment. With regard to the importance of the materials, most of the teachers considered such materials as very useful for their lessons, they could be used as broad guides for future lesson preparations, and they represented an opportunity for them to engage in a learning process while working in their own environment. In relation to the design and tryout of the materials, it is worthy mentioning a number of practical aspects that arose during the processes which needed to be carefully monitored. These aspects include: the role of the teacher in guiding student activities; the role of students as group workers; the time involved in discussions; and the overall monitoring of both teacher and student behaviour as compared to the normal classes. All of these aspects relate to what Black and William (2006) and the Assessment Reform Group (ARG) (1999) from the UK refer to as effective assessment *for* learning, as opposed to assessment *of* learning. ARG (1999) argues that the heart of learning evidence lies in the power of formative assessment and that any feedback for students is only effective if used to guide improvement. In addition, effective assessment for learning depends on effective feedback to students, on their active involvement, and on the adjustment of teaching to the results of student assessment. Harlen (2006) discusses assessment *for* learning as a cycle of events with the students in the centre of it. The cycle

starts with the goals or objectives of the assessment task through which the teachers intend to collect evidence. Then, in possession of enough evidence, it follows an interpretation process aimed at judging the students' achievement so that decisions about next steps can be properly taken. Finally, teachers decide about how to take next steps related to student activities in the learning process, which in turn are directed towards the assessment goals. In fact, one of the issues that the present study wishes to address is related to limited assessment strategies used by Mozambican teachers to collect evidence of learning in schools. The point is that, seemingly, teachers do not only collect insufficient evidence for learning, but also the little evidence being collected, is of poor quality. It seems that teachers apply assessment strategies that allow them to obtain learning evidence only of basic knowledge and skills and even this evidence is not used for improving student learning. The interpretation of the evidence to judge students' achievement is only criterion-referenced, and the ultimate assessment goal is to report on students' achievement. So, one may conclude that the assessment process of Mozambican teachers does not represent Harlen's complete cycle of assessment events and this leads to an ineffective assessment for learning.

Having reviewed some of the intervention studies carried out in the African context and presented the lessons that can be derived from these studies, the next section provides a platform on how these lessons can be used to conceptualise the study and to guide the formulation of preliminary operational research questions of this study.

### **3.6 Summary and conceptualisation of the study**

The topic of this study is to investigate assessment practices used by Grade 12 teachers in Physics in Mozambique and, if needed, to develop an intervention aimed at improving the quality of classroom assessment. Where the literature was reviewed from various angles, the findings were summarised from the perspective of the research questions. This section summarises what was learnt from the reviewed literature as a whole and provides direction about the conceptualisation of the study.



To begin with, one of the most relevant theories in students' knowledge construction is constructivism (Mutimucuo, 1998; Treagust et al., 1996). It was argued that it represents a powerful theoretical resource that may maximise student learning. In fact both the current Mozambican Grade 11 and 12 syllabuses for Physics, and the secondary school curriculum under review, acknowledge and recommend its utilisation. The problem, however, is that while it is recommended for the students in schools, training institutions are still educating teachers within the paradigm of behaviourism. The challenge of this study is to improve assessment practices of teachers and ultimately to help them implement the recommended curriculum.

In the second place, depending on the assessment objectives to be achieved and the context in which the assessment takes place, student achievement in Physics learning should be assessed using different assessment strategies and in varied learning contexts. Therefore, in the process of investigating assessment practices being used by secondary school teachers, there is a need to be constantly alert to what the teacher actually needs to achieve taking into account the conditions in which she or he is working.

In the third place, there is the crucial role played by both teachers and students in assessment. Teachers must know why to assess, what to assess, and understand the importance of quality feedback which they provide to their students. Successful assessment practices take place with active involvement of the students, their ability to assess their peers and themselves, and on their motivation and self esteem. This is likely to occur when using those varied assessment practices that emphasise formative approaches.

In the fourth place, one of the most successful assessment practices in Science education is performance assessment because of its crucial role in assessing student's day-to-day activities. This type of assessment calls upon the students to demonstrate specific skills and competencies and requires them to perform real-world tasks that demonstrate meaningful application of essential skills and knowledge. So, the improvement of teacher assessment practices sought by this study, as will be argued in Chapter 5 (Section 5.3), implies taking into consideration the importance of performance assessment without,

however, neglecting the role played by all other different assessment strategies described in subsection 3.4.4.

In the fifth place, the reviewed literature has put an emphasis on assessment *for* learning, where the results are used to inform the teaching and learning process, as opposed to assessment *of* learning, which is mainly for grading and certification. In undertaking assessment *for* learning teachers must consider completing the entire cycle of assessment events if it is to enhance learning. This cycle of assessment should:

1. Determine the goals of learning and therefore of assessment.
2. Collect enough evidence of learning.
3. Judge whether students' achievement is sufficient.
4. Decide on the next steps in the process of learning and teaching.

Harlen (2006) in this respect, points out that if assessment is to be effective for learning, an entire cycle of goals-evidence-judgment of achievement-next steps in learning-goals has to be completed. The teacher must collect evidence related to goals; interpret the evidence in order to judge the student's achievement; use achievement data to influence decisions about the next steps in learning geared towards the goals. However, and according to some literature reviewed (see, for instance, INDE, 2005; Lauchande, 2001), it seems that Mozambican teachers do not follow the complete process of conducting an effective assessment to inform learning. Teachers do not seem to collect enough evidence of learning - due in part to the use of limited assessment strategies - and they do not use the information they collect to help students learn either. Therefore, the teachers do not complete the cycle of events that might characterise an effective assessment for learning.

Arguments from literature indicate that teachers must be helped to put assessment into context by considering aspects such as why to assess, what to assess, what quality of feedback they should provide to their students, and the curriculum perspective. In this respect, criterion-referenced assessment must be the basis for judging the student performance, the feedback must be judged and used by both students and teachers, and the assessment in general should be directed *for* learning. From this lesson, it is suggested that

this study addresses the issue of ‘collecting evidence for learning’ and its interpretation for judging the student achievement. Furthermore, formative assessment should be considered as the heart of learning evidence and, as supported by ARG (1999), feedback for students will only be effective if it is used to guide improvement.

A concluding remark is that, although all these authors stress the importance of using the collected evidence to take decisions about the next steps in learning, the literature review has shown that somehow there is neither sufficient research of the extent to which assessment strategies are being used for Physics as a subject, nor any reported professional support for teachers to assist them in the development of performance assessment material for use in an ordinary classroom environment. This means that the main question posed by this study remains unanswered in the review of the literature.

This study addresses these shortcomings by contextualising the problem with a focus on secondary school Physics for Grade 12. Apart from the constructivist approach, three other lessons learnt from the literature review can be highlighted as far as the characteristics of materials are concerned. Firstly, there is the need to help teachers by developing and letting them use exemplary support materials on performance assessments that can help students construct their knowledge. Exemplary materials of this nature should help teachers in several aspects of subject knowledge, lesson preparation, teaching methodology, and assessment and feedback. These characteristics should guide them for future lesson preparations, provide them with the opportunity to learn while working, and help them to learn how to develop the materials for topics other than the ones selected for this study. The design and tryout of the materials aspects such as the role of the teacher, the role of students, the class management, and the overall monitoring of student behaviour during lessons, are also aspects to feature in the materials. Secondly, the development of such materials should be done in an ordinary classroom environment to allow users to participate in the process while working in their normal routine. Thirdly the learning evidence should be used to feed the teaching and learning process and, hence, formative assessment is crucial.

In this study the following main research question was examined by the intervention:

*How can the teacher assessment practices be improved?*

In order to address this question adequately, and drawing from the literature, it is of paramount importance to design an intervention that builds upon teachers' present knowledge, skills and experiences with formative assessment in the classroom. As referred to in previous chapters, this implies a prior investigation using a survey approach and aimed at knowing what assessment practices Grade 12 teachers in Physics in Mozambique apply. Some operational research questions are formulated for the Baseline Survey, and are listed below.

- *What assessments practices do Grade 12 teachers apply?*
- *What can be said about the quality of the assessment practices?*
- *How relevant are the assessment practices for student learning?*

Although these research questions generate valuable baseline knowledge about the actual classroom assessment, this knowledge is descriptive by nature and does not provide indications as how to improve the teacher assessment practices as implied by the main research question. The improvement of teacher assessment practices is achieved through working together with teachers in producing and using assessment materials. Briefly, and as was referred to in Chapter 1 (Section 1.1), the main question is twofold, i.e., it implies knowing firstly, what assessment practices Grade 12 teachers apply (Baseline Survey) and secondly how to improve them (Intervention Study).

All baseline and intervention research questions are described in detail in Chapter 4 (Research Design and Methods). At this point, it is relevant to capture what the reviewed literature says about how to improve teacher assessment practices, and what are the most common assessment practices used to assess Science learning.

From the several assessment practices recommended by the literature and from which an intervention for improvement can be conducted, a choice is made concerning performance type of assessment as the focus of this Intervention Study. As was referred to earlier, the rationale for this choice is that, amongst all mentioned assessment strategies, performance assessments appear to be of vital importance in assessing the student understanding of key physical concepts. Performance assessment can also be seen as an adequate means of improving teacher practices of assessing physical and related skills, because all schools expect students to demonstrate a number of skills, from simple communications skills like reading, writing, and speaking, to more complex psychomotor skills like building a car-toy or setting up laboratory equipment. In spite of the importance of all these student skills, when it comes to Science subjects like Physics, the students must not only be able to grasp the concept or the process, but also to explain and use it to solve real-life problems. For example, after students have learnt to identify the direction of power in one electric circuit (e.g., via multiple-choice tests), they must be able to go through the process of identifying, by themselves, some other unknown directions of electric circuits given to them. This kind of hands-on demonstrations of concept mastery is essential in Physics.

In this regard, Airasian (2000) argues that there has indeed been growing emphasis on using performance assessment to determine student understanding of the concepts they are taught and to measure their ability to apply procedural knowledge. Gronlund (1998) also emphasises the role of performance assessment in providing a systematic way of evaluating reasoning and skill outcomes. These outcomes are important, for instance, for Physics because the subject is concerned with solving problems and developing laboratory skills. Moreover, the current educational trend to shift from norm-referenced assessment (ranking of students in order of achievement) to criterion-referenced assessment (description of what students can do) has created a need for a more direct assessment of how well students can perform. Therefore, it is important for this study to allow students to demonstrate, through performance assessment, their ability to do real-world tasks while observing all the procedures involved. It is also relevant to emphasise that, while all other assessment strategies can successfully be conducted in a classroom environment or as homework, effective performance assessment is most likely to succeed when: (i) it is

undertaken in a laboratory context where students can perform real demonstration experiments; (ii) a set of procedural steps is followed – ranging from specifying clear performance outcomes to selecting a proper method of observing, recording and scoring; and (iii) a systematic method of combining them with traditional tests is used. The specific use of any of the above-mentioned assessment strategies depends on the specific learning outcomes to be achieved. This means that to select an adequate assessment strategy, a number of intended learning outcomes must be prespecified. Effectively, for this study the performance outcomes have been identified as the need to demonstrate and develop explanations about force and inertia. Each of these outcomes corresponds to certain areas of performance being assessed. Performance outcomes commonly use verbs such as ‘identify’, ‘construct’, ‘demonstrate’ or appropriate synonyms.

In relation to what aspects of teacher assessment can be taken into account in order to know what assessment practices Grade 12 teachers actually do apply in a contextualised environment, the literature emphasises (as already has been concluded), amongst other formative assessment practices, the importance of performance assessment. However, it is also a lesson from the literature that a pre-requisite for students to be able to learn Physics better, is the prior understanding of the basic concepts related to the subject and, according to the literature, this can normally be assessed using mainly paper-and-pencil tests. Among the variety of other recommended formative assessment practices are observation methods, oral questioning, peer-assessment and portfolios.

These and other assessment practices are examined by the Baseline Survey reported in Chapter 5, while Chapter 4 describes procedural steps, approach, learning outcomes, and performance areas of the Intervention Study. Specifically, Chapter 4 presents the research design of the study (as a rationale for having two phases); the operational research questions of each component (following from this preliminary formulation); the research paradigm; and all methodological aspects of the two phases.

# **CHAPTER 4**

## **RESEARCH DESIGN AND METHODS**

## CHAPTER 4

### RESEARCH DESIGN AND METHODS

*This chapter introduces the research design and methods of the study of investigating and improving assessment practices of Grade 12 Physics teachers in Mozambique. Section 4.1 presents the rationale for having two phases, the research approach for each phase, the general formulation of the research questions addressed by the phases and some reflections on research methodology. Section 4.2 discusses the research paradigm, which was chosen from the research questions addressed in the previous subsection. Section 4.3 elaborates on the research design of the study. The section starts by presenting the research design of the Baseline Survey (the first phase of the study). The research questions addressed by this phase of the study, population and sampling, data collection strategies, and data processing and analysis methods are also presented in this part. The section also discusses the research design of the Intervention Study (second phase), the educational design research as the approach followed in this phase, and the guidelines for designing the intervention. Section 4.4 presents arguments about the validity and reliability of the study while ethical issues are discussed in Section 4.5. Finally, Section 4.6 presents the conclusion and provides an orientation for the following chapter.*

#### 4.1 Introduction

The purpose of this study was to investigate the assessment practices of Grade 12 Physics teachers in Mozambique and how these practices can be improved. The rationale for this study, as discussed in Chapter 1, was that the quality of Physics learning demonstrated by students leaving secondary school is poor and there are reasons for believing that inadequate assessment practices are one of the main contributory reasons for this. As was referred to in Chapter 2, the problem was perceived as a problem at school level. Therefore, it was essential to have a good understanding of the present assessment



practices carried out by secondary school teachers in schools and classrooms in order to design an effective ‘intervention in assessment’. The context in schools can be characterised by various influences from different educational and social entities (see Chapter 2, Section 2.3). In order to gather relevant information pertaining to assessment practices in such a diversified target population, a study by means of a variety of data collection strategies had to be undertaken so that findings can reflect the characteristics of the wider population. This implied that this research should have a preliminary Baseline Survey to develop a good understanding and insight prior to the Intervention Study aimed at designing an intervention that included developing Physics assessment prototypes for teachers to use in their classrooms to optimise the teaching and learning of Grade 12 Physics in the classroom.

The Baseline Survey focused on the identification of assessment practices currently used by Grade 12 teachers in Mozambican schools and their knowledge and skills on assessing students. Teacher knowledge and skills are addressed by investigating the quality and relevance of the classroom assessments. The main research question for the Baseline Survey was formulated as follows: *What assessment practices do Grade 12 teachers in Physics in Mozambique apply and what is their quality?* This research question is in line with the aim of the study which is divided into three specific research topics namely (i) the types of assessment practices (diagnostic, formative, summative) currently in use by Grade 12 Physics teachers in schools, (ii) the quality of these practices and (iii) their relevance for classroom practice. Therefore, the research question is also operationalised into three operational research questions, which are:

1. *What assessments practices do Grade 12 teachers apply?*
2. *What can be said about the quality of the assessment practices?*
3. *How relevant are the assessment practices for student learning?*

The design of the survey is based on the context in which Physics teachers are working in schools as well as on the insights of what the literature highlights as good practice in classroom assessment. More generally, the Baseline Survey lays down the groundwork for

the Intervention Study. Based on the assumption that the assessment practices will help teachers to develop abilities to monitor improvements in student learning and in the performance of the educational system, questions about the types of assessment practices, their quality, and their relevance for classroom practice were included in the Baseline Survey. This assumption was supported by views from the literature reviewed in Chapter 3 (subsection 3.4.4), other authors (Black et al., 2003; Popham, 2002; Weeden et al., 2002), and also from the Science education experts in general.

In light of the findings from the Baseline Survey, the main research question of the Intervention Study – which became the main research question of the study - is: *How can teacher assessment practices be improved?* The process of reviewing the literature on the importance of improving teacher assessment practices (Chapter 3, Section 3.5) has emphasised the need to support teachers both in terms of conducting effective assessment *for learning*, where the assessment results are used to enhance the teaching and learning, as well as the development of authentic assessment material in a classroom environment. According to van den Akker (1999), an evolutionary prototyping of curricular or assessment products and their subsequent representations in practice are viewed as more productive than linear development approaches. Formative evaluations of subsequent assessment versions are essential to such productiveness, and an educational design research approach is seen to enhance knowledge growth. Therefore this intervention, as discussed in Chapter 3 (Section 3.6), focuses on (i) the design and formative evaluation of exemplary performance assessment materials for demonstration experiments aimed at assisting teachers to improve their assessment practices and on (ii) a laboratory written report by students. A great emphasis has been put on lesson materials rather than on assessment materials throughout the intervention. This is because any assessment strategy can only be successful if it is applied using quality lesson materials. Good lesson materials provide adequate support information for the preparation of student assessments, they are broad guides for future lesson preparations (including assessment tasks), and teachers and students implement most of their lesson activities according to advice provided in the lesson materials.

The demonstration experiments and the students' report are designed to focus on only two Physics concepts namely force and inertia, whereas the intervention addresses the functions of assessment namely diagnostic, formative and summative assessment. The reasons for selecting these two Physics concepts are given in subsection 4.4.3. As stated earlier in Chapter 3 (Section 3.1) 'demonstration experiments' refers to students' activities of observing, carrying out small experiments, interpreting phenomena, and reporting findings, and are all guided by a teacher. Demonstration experiments may be performed either individually or in small groups of students and they take a few minutes to perform. When these experiments take longer to perform (from 30 min to hours), the literature refers to them as laboratory demonstration experiments. In this study, a decision was made to use the term 'demonstration experiments' because of the characteristics described above.

The intervention applies the methodological approach of educational design research suggested by van den Akker and Plomp (1993). The potential of educational design research is that the search for characteristics of an effective intervention is conducted while working on that intervention. The research approach is discussed in Section 4.3 while the research paradigm, considered suitable for addressing the research questions of the study, is discussed in the next section, Section 4.2.

## **4.2 Research paradigm**

Several authors have argued that to choose the type of knowledge claim, the researchers have to adapt certain assumptions about what and how they will learn during their inquiry (Creswell, 2003; Lincoln & Guba, 2000; Mertens, 1998). According to these authors, this claim can be named ontology, epistemology, philosophical assumption, or paradigm. Philosophically, the researcher makes claims about the nature of the reality, i.e., what is knowable (ontology), what is the relationship between the researcher and the researched (epistemology), the language of the research (rhetoric), and what the process of studying the reality (methodology) will be.

According to Creswell (2003), there are four schools of thought about knowledge claims namely post-positivism, constructivism, advocacy, and pragmatism. Post-positivist researchers claim that causes probably determine effects. These researchers challenge the traditional notion of the absolute truth of knowledge and they claim that when studying the behaviour of human beings one cannot be positive about the claims of knowledge. Constructivists often address the process of interaction among individuals and focus on the specific contexts in which they live and work in order to understand their historical and cultural settings. Advocacy researchers believe that inquiry needs to be intertwined with politics, and the research should contain an action agenda that may change the lives of the researched, the institutions where they work, and the researcher's life. Pragmatists claim that knowledge arises out of actions, situations and consequences rather than antecedent conditions, and their main concern is with applications and solutions to problems.

Within this framework, the scientific position of this study (as referred to in Chapter 1, Section 1.3) is rooted in the pragmatic knowledge claim. The research process did not begin with any one system of reality to identify the type of research method to be applied; rather, it started by identifying the problems to be solved, i.e., from the research questions formulated and went on to identify the suitable research methods that were relevant in obtaining valid and reliable answers to these questions. The research was geared towards the best understanding of the research problem. The truth was not solely based on dualism between the researcher's mind and reality, but it was on what worked at the time. Both qualitative and quantitative methods were applied to collect and analyse data with the main aim of understanding the complexities of the current situation and to produce findings that contribute to a solution to the problem.

### **4.3 Overview of the research design**

In line with the pragmatist claim described above, this study intended firstly, to find out what classroom assessment practices teachers are using in schools before choosing any particular type of assessment to monitor improvements and secondly, to generate a methodological approach and guidelines for the design and development of an adequate

study approach aimed at improving such practices. The most suitable strategy to identify current assessment practices is the survey approach, where a study was conducted using questionnaires, observations, and interviews for collecting data. As several authors quoted in Chapter 3 have argued (Airasian, 2001; Dekkers, 1997; Moskal, 2003; Stiggins, 1987), in order to expand teacher assessment practices in Science subjects, demonstration experiments are important to improve performance assessments, particularly in Physics (White & Gunstone, 1992). The intervention was aimed at improving teacher assessment practices in Physics, and was conducted in such a way that teachers and students could conduct demonstration experiments performing real-world tasks while working in their normal classroom schedule under existing conditions and materials. Thus, the improvement of assessment practices proposed by this study was investigated under ordinary classroom circumstances and not in a setting specifically created for this research. So, the study approach was geared towards what works in schools and how this can be improved on the basis of intended consequences.

Methodologically the study applied, for the survey approach, mixed methods in recognition of the fact that both quantitative and qualitative methods may have limitations and one can neutralise the limitations and biases of the other (Creswell, 2003). Triangulation was considered as a means to seek convergence of findings. Still regarding the survey, the principle of using different data sources and multiple data collection instruments was used to guarantee triangulation. For the intervention, the strategy consisted of formative evaluations of exemplary assessment materials (specifically designed for this study) where the quality was verified by investigating the validity, expected practicality and expected effectiveness of the materials produced (Nieveen, 1997; van den Akker, 1999; van den Akker and Plomp, 1993).

Subsection 4.3.1 presents the research design of the Baseline Survey. It discusses the methodological aspects of operational research questions, population and sampling, instrumentation and data analysis.

### ***4.3.1 Research design for the Baseline Survey***

In order to address the objective of the research, it was necessary to start by undertaking a preliminary identification of assessment practices currently used by Grade 12 Physics teachers in Mozambican schools. Specifically, this implies a search of the kind of assessment practices Grade 12 teachers of Physics currently apply at classroom level and what can be said about the quality of these practices. The country has a population of 120 secondary school teachers teaching Physics in Grade 12, distributed in 30 schools of General Secondary Education – cycle 2 (ESG2) (per June 2004). As the purpose of the Baseline Survey was to inform the Intervention Study about assessment practices used in this target population of teachers and schools, and not to gain a national representative picture, it was believed that a small survey of some purposively selected Mozambican secondary schools from different provinces would be sufficient. In other words, it has been assumed that the perspective of various purposively selected school contexts would be representative for the characteristics of teachers, students, and schools.

As indicated in Chapter 3 (Section 3.6), a preliminary research aimed at identifying the assessment practices used by secondary school Physics teachers in Mozambique, was undertaken. The aim is expressed in three operational research questions. Because the research questions are formulated in line with the three corresponding research purposes (see Section 4.1), it is useful and relevant to gain a good understanding of a number of characteristics of the assessment practices applied by teachers, viz. the types of assessment practices, their quality, and their relevance for learning. These three elements constitute the perspective from which the characteristics of assessment practices are viewed during the survey in schools.

The identification of types of teacher assessment practices to be looked at in the classroom was firstly informed by the literature review and later refined by the pilot phase of the data collection instruments. From the variety of formative assessment practices referred to by the literature as crucial for what teachers need to know in order to undertake a contextualised assessment, observation methods, oral questioning, peer-assessment, and portfolios are the most critical (see Chapter 3, subsection 3.4.3). While oral questioning,

peer-assessment and portfolios were directly observed from the classroom by the researcher, teachers' own observation methods were not easily observable. In order to accomplish this, the strategy was to analyse how teachers observed students' process of designing and developing finished products resulting from a certain planned activity. From this analysis, it was possible to record the teacher's own comments and suggestions for improvement. Finished products include all written tasks that the students do and which reflect a certain development process of collecting, interpreting and reporting data. These products are known in schools as projects. In addition to these assessment practices, paper-and-pencil tests were also investigated due to their potential in assessing student abilities to understand basic concepts.

As a result of lessons learnt from literature and improvements arising from the pilot phase the following types of teacher assessment practices used in the classrooms, namely portfolios, peer-assessment, verbal tests, paper-and-pencil tests and projects were investigated in this study. These assessment practices are deemed relevant and good types of assessment to be used in classroom assessments (Popham, 2002; Weeden et al., 2002). Since these practices had been already referred to by the literature as assessment strategies that teachers normally use in schools, the identification of the types for this study was done by verifying which ones were more frequently used by Mozambican secondary school teachers to assess their students.

The term 'quality' of assessment practices refers to all aspects of validity and reliability of these practices. As referred to earlier in Section 4.1, the quality of assessment practices includes the teacher knowledge and skills for assessing student work. Thus, the quality aspect was investigated by analysing how teachers were assessing oral communication during lessons, written work, presentations, notebooks, and laboratory work of their students. According to several authors, these student tasks, if regularly undertaken, are indicators of the quality of assessment, particularly for science subjects (Black et al., 2003; Race et al., 2005; Weeden et al., 2002). In the context of this study, the quality of assessment practices used by teachers was investigated by two means: (i) verifying the

frequency of use, by teachers, of these student tasks and (ii) checking their validity and reliability.

As in the two previous aspects, the element of *relevance* was investigated on the basis of what the literature suggests is good practice. Popham (2002) for instance, argues that if the teacher shares the goals to be achieved with the students and involves them in the evaluation of their own work, this allows students to know what is expected from them. Thus, the issue of assessment relevance was addressed by investigating (i) how teachers engage students in the evaluation of their performance, and (ii) how often they use the assessment results to guide the student learning.

This section is composed of four subsections. Subsection 4.3.1.1 elaborates on the three operational research questions referred to earlier in this chapter and how the research design will address these. Subsection 4.3.1.2 presents the population and the sample of the Baseline Survey. Subsection 4.3.1.3 describes the instrument development. The process of instrument development and piloting, triangulation, identification of data sources, as well as the procedures followed during the survey, are all presented and discussed in this subsection. Subsection 4.3.1.4 presents the methods used for analysing data.

#### ***4.3.1.1 Research questions for the Baseline Survey***

The main research question and the three operational research questions are presented and described in this subsection. The main research question addressed by the Baseline Survey is:

**What assessment practices do Grade 12 teachers in Physics in Mozambique apply and what is their quality and relevance?**

This question was addressed by a survey of 12 teachers from six schools in Gaza, Zambézia and Cabo Delgado Provinces, and of five educational officers from the Ministry of Education and Culture in Maputo. Taking into account the aspects of which



characteristics of assessment practices were to be investigated (types, quality and relevance), the formulation of the operational research questions for the Baseline Survey followed the same classification. Three operational research questions were formulated. The first operational research question sought to identify the types of assessment practices used by teachers, namely:

***a) What assessment practices do Grade 12 Physics teachers apply?***

Assessment practices investigated in the classroom included the five types mentioned earlier, namely portfolios, peer-assessment, verbal tests, paper-and-pencil tests, and projects. As was mentioned earlier, the criterion used to identify the types of assessment practices used by teachers in schools was to verify and count how many times (i.e., how often) the teachers used each assessment practice during several classroom sessions. Therefore, in order to address this question the teachers were asked to give information, amongst other questions, on the following sub-question:

- *a. How often do you use each of the following assessment practices? Portfolios, peer-assessment, verbal tests, paper-and-pencil tests, projects*

By means of questionnaires, interviews and through classroom observations, it was possible to verify - by checking the frequency (daily, weekly, monthly, never) of using the different assessment practices - which assessment practices the teachers have been applied during the classroom assessment. Teachers were also allowed to describe other possible assessment practices, which they use.

The second operational research question addresses the quality of the assessment, namely:

***b) What is the quality of the assessment practices?***

Aspects of assessment quality include not only the frequency but also the characteristics of the assessment tasks in terms of students' knowledge or skills (reasoning, memory or

process) being assessed by certain assessment type or activity. These elements were addressed by the following sub-questions:

- *b1. How often do you assess the following student activities? Oral communication during lessons, written work, presentations, exercise books, laboratory work, solving problems*
- *b2. What can be said about the validity and reliability of the assessment practices?*

The validity and reliability of the student activities was verified by the kind of feedback given to students by teachers in those different activities. Aspects of expression (whether the feedback was congratulatory or critical), time (timely feedback or given afterwards), and personality (individualised or in-group feedback) were used for the most assessed activities (Race et al., 2005).

Finally, the third operational research question deals with the relevance of these assessment practices for learning, namely:

***c) How relevant are the assessment practices for student learning?***

The relevance of the assessment practices refers to those elements that express the level of students' involvement in their own assessment, as well as the follow-up actions to be undertaken by the teacher after handing the assessment results out. Two sub-questions were formulated to address this question, namely:

- *c1. How do you engage students in the evaluation of their performance?*
- *c2. How often do you use the assessment results, for what purposes, and how?*

Teachers were given several alternative options on students' involvement in the evaluation of their performance namely (i) I do not involve them at all; (ii) by handing the results out; (iii) by involving them in self-assessment; (iv) by sharing with them the goals to be achieved; (v) by explaining to them the implications of the evaluation; (vi) by reflecting with them on the assessment data. Particular emphasis was given to peer-assessment due to the impact of this type of assessment in self-assessment (Race et al., 2005).

The three operational research questions were investigated using various target populations, which are described in the following section.

#### ***4.3.1.2 Population and sampling***

There were three target populations relevant for addressing the operational research questions of the Baseline Survey. These are listed below.

- Teachers – are the active subjects in the assessment processes being investigated.
- School directors - have the responsibility for implementing the government regulations on assessment and on monitoring the quality of teaching in their school. They also play a role in creating a supportive school culture.
- Education officers – are responsible for providing the infrastructure to schools and inspect whether schools do a good job in terms of quality education.

As the Baseline Survey was aimed at gaining an impression of the assessment practices of Physics teachers in schools, it was believed that a small survey of Physics teachers and school directors from six Mozambican secondary schools from various provinces, representing the different contexts (urban-rural, different regions in the country), would be sufficient for this purpose (see also Chapter 1, Section 1.3 and Chapter 4, subsection 4.3.1). As referred to in Chapter 2 (Section 2.1), the country is composed of eleven provinces, clustered into the North, Centre, and South. Three provinces were drawn from these regions – one from each – and two schools were selected from each of these provinces. Maputo City was only considered for pedagogical officers and assessment specialists. No schools were selected from Maputo City because schools in this area appeared to be exhausted by extensive research activities taking place at the time. In order to enable comparison between teachers' responses, schools were selected according to their capacity of having at least two teachers teaching Physics in Grade 12. However, there were still three schools in which only one teacher taught Grade 12 Physics. This was the case in Pemba and Montepuez Secondary Schools in Cabo Delgado and Mocuba

Secondary School in Zambézia. Where this occurred, an additional teacher was taken from Grade 11. In the end, two Physics teachers from each of the two selected schools in the three provinces (in total 12 teachers) were sampled. The school directors were also sampled for participation in the study. In total, the intended sample was composed of twelve teachers and six school directors. Only four school directors, however, participated in the study as such, because two of the school directors were also Physics teachers and, due to practical limitations, they could only participate in one capacity. Given the focus of the study the role of the teacher was considered more important and, therefore, they had to provide information as teachers and not as directors. It was very important to obtain as much information as possible about the assessment practices carried out by teachers and the number of teachers excluding these ones would have been insufficient for this purpose.

Besides, five educational officers (two pedagogical officers and three assessment specialists) from the Ministry of Education and Culture (MEC) were asked to participate in the study. The pedagogical officers and assessment specialists were purposefully selected from the MEC in Maputo City due to their responsibilities for monitoring the assessment system within the Ministry. Of these pedagogical officers, one is the Director of the National Institute for Educational Development (INDE) - an institution responsible for curriculum review for both primary and secondary education - and a former Head of the Department of Assessment and Certification in the Ministry, and the other is the National Education Inspector. Concerning the assessment specialists, all of them were science subject specialists working in different departments within the Ministry. Table 4.1 summarises the details of the realised sample for the Baseline Study.

**Table 4.1: Sample of Baseline Survey**

Region	Province	Institution	Nr. of teachers	Nr. of school directors	Nr. of educational officers
South	Gaza	Joaquim Chissano Secondary School	2	1	-
		Chókwè Secondary School	2	1	-
	Maputo City	Ministry of Education and Culture	-	-	5
Centre	Zambézia	25 de Setembro Secondary School	2	1	-
		Mocuba Secondary School	2	-	-
North	Cabo Delgado	Pemba Secondary School	2	1	-
		Montepuez Secondary School	2	-	-
<b>Total</b>	<b>3</b>	<b>7</b>	<b>12</b>	<b>4</b>	<b>5</b>

The samples of schools (and therefore the samples of teachers, school directors and educational (both pedagogical and assessment) officers were *purposive samples*, as they were drawn with the purpose of obtaining insight into three important perspectives on the classroom practice namely instruction in relation to assessment, management’s perspective in relation to teachers’ preparedness for conducting appropriate assessments, and inspectorate regarding quality control of the teacher assessment practices, with the view to using the information to design an intervention study and not to generalise to full populations. With these samples, all activities were undertaken to address the main research question for the Baseline Survey formulated at the beginning of subsection 4.3.1.1.

#### 4.3.1.3 Data collection strategies

This subsection comprises four parts. The first part presents the number and characteristics of data collection instruments used in this study, and a summary of the content of each instrument. The second part discusses the development process of the various instruments, the piloting process, and the validation of the instruments by experts. The third part starts by providing information on the type of data collected to answer operational research questions, the way it was collected, the triangulation process of instruments and data sources, and ends with a summary of all information in a data

collection matrix. The fourth part presents procedural information on the number and sequence of activities that were carried out in preparing and conducting the Baseline Survey for this research.

#### *Instruments and data collection strategies*

Five data collection instruments were developed for the Baseline Survey namely, a questionnaire for teachers, a classroom observation schedule, and three interview schedules for teachers, school directors, and pedagogical officers.

The questionnaire for teachers (Appendix A) consisted of four main sections. The first section contained information about the questionnaire itself (e.g., what is it about, why should it be filled in) and requested personal background information about the respondents (name, gender, age, school, etc.). The second section requested information about the types and quality of assessment practices used by teachers in the classroom. This information was sought through five closed questions of multiple choice and Likert scale type of items. The third section was about the relevance of the assessment practices and comprised of four questions of multiple choice items. The fourth and final section contained evaluation questions with one question containing multiple-choice items and another one being open-ended. It is worthy to mention that, although the majority of the questions were closed, they provided teachers with the opportunity to express their views and opinions by exploring the ‘other specify\_\_’ type of items. Furthermore, in the evaluation questions teachers were asked to comment about any other issue that was not addressed in the questionnaire.

The classroom observation schedule (Appendix B) contained four sections. The first section contained background information about the teacher and the school, the second was about the physical appearance of the classroom and the teaching and learning environment. The third section presented the description of the students (e.g., number, gender, age), and the fourth section comprised a number of close questions (Likert scale items) related to the types of assessment practices undertaken by the teachers and their relevance for learning. More specifically, the questions addressed the extent to which

assessment practices were applied by teachers, their quality as demonstrated by teachers and students, their appropriateness for instruction, and their validity and reliability for student learning.

The interview schedules for teachers, school directors, and pedagogical officers (Appendices C, D, and E) comprised of 13, nine, and ten questions respectively. All schedules had an introduction stating the aim of the interview and the reasons why the interviewees and their respective schools were chosen to participate in the study. The introduction also indicated that the identity of the interviewees would remain anonymous and the information confidential. The interview questions addressed similar issues as in the questionnaires such as types, quality and relevance of assessment practices with the intent to cross check the information. There were three additional aspects addressed in the interviews with the pedagogical officers from the MEC. The first aspect was related to the objectives of the teachers' assessment as seen by the Ministry. The second asked how the translation of these objectives into practice is compared to the information provided by the national examinations undertaken by the Ministry. The third was meant to seek the opinions of the interviewees about the impact of the supervisory visits to schools.

Overall, the questionnaires and interviews were designed to gather information about the types, quality and relevance of assessment practices used by secondary school teachers in schools. The classroom observation schedule was also used by the researcher to triangulate the information on assessment practices given by teachers and school directors. But more importantly, it permitted the observation of the physical conditions of the classrooms, the teaching and learning environment, and the characteristics of the students. The different instruments used with different data sources allowed for cross checking of the information and increased its validity and reliability.

#### *Development process and piloting*

The first version of data collection instruments for Baseline Survey was developed by the researcher and the instruments were appraised by experts to ensure their validity. The instruments were all piloted before the data collection process. Questionnaires and

interview schedules for teachers and school directors, as well as, classroom observation schedules, were piloted with five Grade 12 Physics teachers randomly selected from schools around Maputo. School directors' interviews were piloted with two directors of some secondary schools also located in the Maputo area. The interviews for pedagogical officers were piloted with two pedagogical officers from INDE. The main objective of the pilot phase was to increase the validity of the instruments in terms of language, depth of assessment content approach, and time required for completing the instruments. For instance, one expert on designing assessment instruments and one Science educator specialist were asked to provide their comments and suggestions on how to improve the quality of the instruments (content validity). Both experts scrutinised all the instruments in order to determine their validity in terms of face validity. The reliability of the instruments was checked by verifying the consistency of the responses and to ensure that respondents answered related items in a similar way (internal reliability). More specifically, the piloting was intended to find out whether all members of the sample, especially teachers, would be able to understand the instruments and to complete the questionnaires in time.

After revision of the first version, the instruments were finalised. All the participants in the pilot phase were asked to comment on the content and practicality of the instruments. Overall, the pilot phase was instrumental in improving the validity and practicality of the data collection instruments by generating valuable suggestions for improving of the final version.

#### *Data collection and triangulation*

As was discussed earlier, in order to gather the information needed to answer the main research question of the Baseline Survey and to assure the validity and reliability of the information gathered the principle of triangulation of data sources and of instruments was applied.

The information needed to obtain answers for the *specific research question a.* included (x1) the types and frequency of usage of a certain type of assessment practice, (x2) the opinions about why teachers assess and evaluate the student performance, including about



the teacher preparedness level and (x3) the physical conditions under which the assessment practices are carried out. This information could best be obtained from questionnaires with the purposively selected teachers. More generally, teachers were asked to reflect about the different assessment practices they use in their classroom, their purposes, how they conduct them in practice, how they evaluate the final performance of the students, and under what physical conditions they work. To validate the information from the teacher questionnaires a number of interviews were conducted with the teachers. In the interviews they also had to respond to the questions about types and quality of assessments with the aim of supporting or refuting their arguments expressed during questionnaires. Classes were observed and interviews conducted with school directors (who are also teachers) and with pedagogical officers for the same purpose.

For the answers to the specific research question b. the information needed included (y1) data about the frequency of assessing, by teachers, of certain students' activities, and the teachers' opinions firstly about whether the assessment practices used allow students to demonstrate their performance (y2) and secondly about the characteristics of their scoring procedures of these practices (y3) – whether they are clear, consistent and unbiased. The information about the quality of assessment practices was mainly obtained through questionnaires (to teachers and school directors) and interviews (to school directors and pedagogical officers). Pedagogical officers were also interviewed to gather their views about the level of preparedness of teachers in designing and administering classroom assessments and as sources of information complementary to the information provided by teachers and school directors. The validation of the information from both data collection instruments was done by researcher's observations of the classroom practices and by written notes provided by the assessment specialists.

Regarding the information necessary to address the *specific research question c.* data included (z1) the use by teachers of assessment results to monitor student learning, and (z2) the teacher preparedness to design classroom assessments including the coverage of relevant topics, and student involvement in the evaluation of their own work. This information was obtained from teachers and school directors through questionnaires. To

validate this information two means were used: (i) the first was the interviews with school directors, both to cross check the information provided by other teachers and the observations during lessons (external reliability) as well as to enable teachers to express their own views about assessment practices currently in use in their schools; (ii) the second was written notes from the assessment specialists who were asked to state in writing not only how they perceive the classroom assessment as taking place in the schools, but also the MEC's philosophy of what they consider as desirable assessment practices for schools as well as the level of teacher accomplishment of them particularly as assessment specialists have been playing a major role in the supervision of the teaching quality and in the development of national assessments.

As referred to earlier, apart from interviews, questionnaires and written notes from assessment specialists, the researcher also conducted classroom observations. Eight classroom observations were conducted with eight teachers; the remaining four teachers did not have their classes observed because they were not available at the time of observations. Classroom observations were deliberately not announced in advance to avoid simulation of lessons. This was also the reason why only one class was observed per teacher where the objective was mainly to obtain information about the unplanned assessment practices, the teaching strategies, and the physical conditions of the typical Physics classroom. Verifying teachers' formal assessment practices was not necessarily the objective of the observations, because formal assessments are planned and announced in advance and they did not take place on the dates of the observations.

Overall, while teachers mainly provided information about type, quality, and frequency of assessment practices, school directors, pedagogical officers and assessment specialists were asked to give their opinions about the quality of assessment practices, the use of assessment results for monitor student learning, and the level of teachers' preparedness in designing acceptable assessment practices.

Thus, in order to guarantee valid and reliable information and for triangulation purposes, a variety of data collection strategies, instruments, and data sources were used to answer the

formulated research questions. This is summarised in the following data collection matrix (Table 4.2).

**Table 4.2: Data collection matrix**

Instrument	From teachers via:			From school directors via:		From pedagogical officers via:	From assessment specialists via:
	Qn	Iv	Ob	Qn	Iv	Iv	Wn
<b>Information variables</b>							
<b>a. Assessment practices used</b>							
-types and frequency (x1)	✓		✓			✓	
-opinions about why do they assess, how much of the teaching time they spend on assessment and how do they evaluate the student performance (x2)		✓			✓	✓	✓
-level of teacher preparedness		✓	✓		✓	✓	
-physical conditions (x3)			✓		✓	✓	
<b>b. Quality of assessment practices</b>							
-frequency of certain student activities (y1)	✓	✓	✓	✓		✓	✓
-allowing students to demonstrate performance (y2)	✓	✓	✓	✓	✓	✓	
-clear, consistent and unbiased scoring procedures, etc. (y2)	✓	✓	✓	✓	✓	✓	
<b>c. Relevance of assessment practices</b>							
- frequency of using student assessment results (z1)	✓	✓		✓	✓	✓	✓
-level of student involvement in evaluating their own work (z2)	✓	✓		✓	✓		✓
-coverage of relevant topics and appropriateness for instruction (z2)	✓	✓		✓	✓		✓

Qn = questionnaire, Iv = interview, Ob = classroom observation, Wn = written notes

In practice, questionnaires and classroom observation schedules were the main data collection instruments. Interviews were conducted after the completion of the

questionnaires and the observation of classes in order to guarantee the reliability of the information. To avoid copying of information, the interviews were conducted on the same day with all teachers, one after another.

### *Research procedures*

The main research activities of this study started with the literature review that was carried out between June 2003 and April 2004. The document analysis at the MEC, aimed at obtaining statistical information about the number of secondary schools teaching Grade 12 Physics in the country as well as their number of Physics teachers, was undertaken from May to June 2004. Then the instrument development phase was initiated where the first version of the self-completed questionnaires, classroom observation schedules, and interview schedules were developed by the researcher. These were then piloted. Then the main fieldwork was conducted during which a number of activities were undertaken sequentially. As was referred to earlier, firstly questionnaires were administered to the twelve selected teachers from the six schools. Then interviews were conducted with teachers and thereafter, classroom observations were undertaken, focusing on teachers' instructional practices. These took place in August 2004 and March 2005. The interviews with the school directors took place from August to October 2005. In November 2005, the interviews with pedagogical officers from the MEC took place. Finally, the written notes which had been requested by assessment specialists and from the MEC (who were previously asked to provide their thoughts in writing about Physics classroom assessment) were collected in February 2006.

The subsection 4.3.1.4 presents the methods used to analyse the data.

#### ***4.3.1.4 Data processing and analysis methods***

Self-completed questionnaires from the teachers were analysed quantitatively using categorisation of questions and calculation of frequencies (Bardin, 1977) following a set of procedures for describing, synthesising, analysing, and interpreting data (Gay & Airasian, 2003). Frequencies were produced from the analysis and were presented in

graphs and tables. The software package, Statistical Package for the Social Sciences (SPSS - version 8.0), was used to directly capture all quantitative data from different data collection instruments. Data were analysed and presented using frequency tables.

Interviews and classroom observations were analysed qualitatively through summarisation of questions and categorisation of the responses (Miles & Huberman, 1994). Thematic content analysis was employed to analyse the data (Bardin, 1977; Race et al., 2005). Contact summary forms with excerpts for illustration were then filled in to review the interview and classroom observation notes and an overall summary of the main findings was produced. Two main concurrent flows of activity were followed, namely (i) a process of deciding on the meaning of each item of data or set of data by noting similar patterns or explanations and (ii) a process of assembling information that allows to draw conclusions and to take further action. Findings from the Baseline Survey are presented and discussed in Chapter 5.

Subsection 4.3.2 is the research design of the Intervention Study. The section introduces the educational design research approach, elaborates on the design of Physics assessment materials prototypes, and ends with a discussion of design guidelines for the intervention.

### ***4.3.2 Research design for the Intervention Study***

The study on investigating and improving assessment practices in Physics in secondary schools in Mozambique focuses, on designing and developing Physics assessment materials aimed at helping teachers to improve their assessment practices in schools. Following from the main research question of this study, as formulated in Chapter 1 (Section 1.2), the aim of this study is twofold, namely to identify assessment practices used by secondary school Physics teachers in Mozambique and to undertake an intervention aimed at improving these practices. The Baseline Survey described in previous sections of this chapter addressed the first part of the aim. The Intervention Study dealt with in this section addresses the second part of the study aim and its research question is formulated as:

## **How can the teacher assessment practices be improved?**

In order to find an answer to this question, and following from what was referred to in Chapter 3 (Section 3.5), the Intervention Study applies an educational design research methodology as suggested by van den Akker and Plomp (1993). Findings of the Baseline Survey indicated that, although teachers have regularly been using paper-and-pencil tests - which are good for grasping the basic concepts, performance assessment is important for students to be able to perform real-world tasks and Physics cannot be taught and assessed without practical or laboratory work (see Chapter 5). This argument led to the reinforcement of what is already the Mozambican government policy of adopting the constructivist approach to learning and teaching. In fact, constructivist perspective underpinned the approach applied in this study in improving teacher assessment practices. A number of Physics assessment prototypes were designed in the context of demonstration experiments and they were formatively evaluated in classroom tryouts.

This study was conducted following a process of analysing the problem context, carrying out a Baseline Survey, recommending a type of intervention to be made, and designing and formatively evaluating assessment prototypes. This process reflected a nature of two mixed method approaches namely survey and educational design research.

Three subsections comprise this section. Subsection 4.3.2.1 elaborates on the nature of the educational design research approach. Subsection 4.3.2.2 discusses educational design research as applied in this study, that is, in the context of the Physics assessment materials. Subsection 4.3.2.3 presents the design guidelines for the Intervention Study. This subsection also provides orientation elements for the prototyping process as design specifications for the Physics assessment materials.

### ***4.3.2.1 Educational design research***

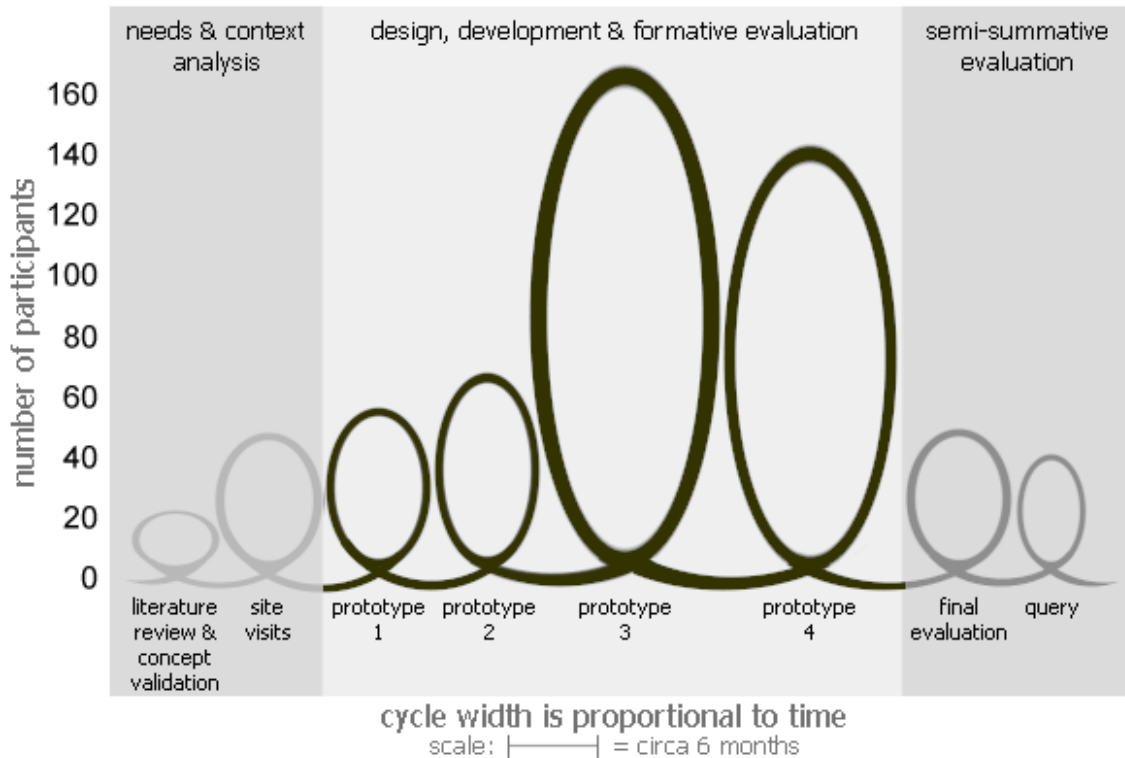
According to van den Akker (1999), educational design research could be an appropriate approach for a complex situation where the appropriateness and effectiveness of an

intervention is unknown beforehand and its success depends on the design and implementation process within a wide variety of the contexts. This is indeed the case for Mozambican schools where many people (particularly teachers, students, and teacher educators) are unfamiliar with the process of designing assessment materials. Furthermore, curriculum materials (including assessment ones) in the country are developed using several theories but rarely utilise findings from research. Therefore, as was referred to in Chapter 2 (subsection 2.2.2), since the Ministry of Education and Culture is in the process of reviewing the curriculum for secondary education (including assessment issues), timely and adequate information is required for the reviewers to make the right choices in such a complex and dynamic situation.

Educational design research is a systematic process of designing, developing and evaluating instructional programs, processes, and products that must meet the criteria of validity, practicality and effectiveness (Seels & Richey, 1994; van den Akker, 1999; van den Akker & Plomp, 1993).

The function of the educational design research, as is the case in any scientific research, is the search for understanding, which results in contributing to the body of knowledge or theory. This search can occur with a very broad purpose through conducting scientific research in a certain domain or at micro level of specific research projects. In education research, the search intends: (i) to contribute to the body of scientific knowledge or theories about education; (ii) to contribute to improving practice; and (iii) to inform decision-making and policy development in the domain of education. Within the context of a research project, various functions of research can be distinguished, namely to describe, to compare, to evaluate, to explain, and to design and develop.

The research process in educational design research (also called development research) comprises educational design processes undertaken in cyclical stages of analysis, design, evaluation and revision activities. The stages could be refined continuously until reaching a satisfying balance between the intended and the realised stage. McKenney (2001) cited by Plomp (2006) gives an illustration of the cyclical process as set out in Figure 4.1.



(Source: McKenney, 2001)

**Figure 4.1: The cyclical process of educational design research**

According to Plomp (2006) three distinctive phases can be distinguished from this example as set out below.

1. Preliminary research: this involves a needs and content analysis, and a literature review (including site visits) leading to a conceptual framework.
2. Prototyping phase: this requires iterative and cyclical design, development with formative evaluation of the several prototypes as the most important research activity aimed at refining the intervention.
3. The Assessment phase involves semi-summative and final evaluation to conclude whether the solution or intervention meets the pre-determined specifications.

Throughout all these activities the researcher will do ‘systematic reflection and documentation’ to produce the theories or design principles as the scientific yield from the research.



As already reported, this study started with the preliminary research on relevant literature about what scholars regard as good practice in classroom assessment, as well as on the context analysis (Baseline Survey) of what assessment practices Grade 12 Physics teachers use in schools.

During the prototyping phase the guiding orientation for employing educational design research in designing exemplary materials and fulfilling the functions mentioned above, is that these prototypes must be of good quality (Nieveen, 1999). Nieveen suggests a framework for making the concept of quality in exemplary materials more transparent, which includes three criteria namely: validity, practicality, and effectiveness. Validity is attained when there is internal consistency between the materials and the state-of-the-art knowledge (content validity), and between the different components of the materials (construct validity). Practicality is attained when the materials are usable by teachers and students in a way that is compatible with the developer's intention. Effectiveness is attained when students appreciate that the desired learning tasks and the learning programme are taking place.

The cyclical character of educational design research does not mean that activities are just undertaken repeatedly. The quality criteria are taken into account and they are given different emphasis in different stages of the research. For example, during the preliminary research where the emphasis is on analysing the problem and reviewing the literature, the criterion of validity is the most dominant, with some attention given to practicality, whilst in that state no note is yet taken of effectiveness. On the other hand, in the prototyping phase much attention is paid in the formative evaluation to the criterion of *practicality*, whilst effectiveness becomes increasingly important in later iterations. Finally, the systematic reflection and documentation is undertaken at the end of each designing cycle, and it is aimed at enhancing design principles and implementation solutions. Table 4.3 depicts the phases of educational design research, the quality criteria emphasised in each phase, and the activities being undertaken.

**Table 4.3: Quality criteria related to phases in educational design research**

	Phase	Criteria	Short description of activities
1	Preliminary research phase	More emphasis on validity, less on practicality	Review of the literature and of (past and/or present) projects addressing questions similar to the ones in this study. This results in a framework and first blueprint for the intervention.
2	Prototyping phase	Initial emphasis on validity and practicality.  <b>Later on</b> mainly practicality and gradually shifted to effectiveness	Development of a sequence of prototypes that will be tried out and revised on the basis of formative evaluations. Early prototypes can be just paper-based for which the formative evaluation takes place via expert judgments.  Evaluate whether target users can work with intervention (practicality) and are willing to apply it in their teaching (relevance & sustainability), also whether the intervention is effective.

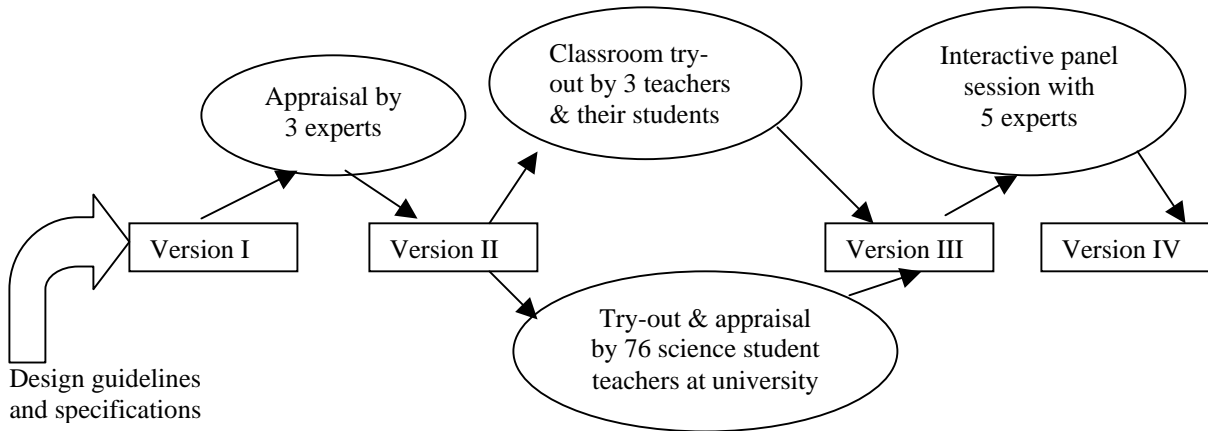
Various terms are used in the literature for the preliminary research activity, such as ‘orientation’ (Hadi, 2002), ‘needs and context analysis’ (McKenney, 2001), ‘front-end-analysis’ (Ottevanger, 2001), and ‘in-depth orientation’ (Thijs, 1999). The preliminary research in the Physics Assessment Materials (PAM) for this study is called ‘Baseline Survey’. In relation to the prototyping activities, some other authors have used different terms such as ‘design and development and evaluation stages’ (McKenney, 2001) and ‘development and evaluation and semi-summative evaluation stages’ (Hadi, 2002), respectively. During the prototyping stage activities consist of tasks to articulate the design ideas into the empirical development stage.

Formative evaluation of the research activities takes place in all phases of educational design research. As illustrated by Table 4.3, it serves different functions in the various development cycles. It also has various layers in design research activities, from the more informal in the early phases of a project (self-evaluation, one-to-one evaluation, expert review) to small group evaluation aimed at testing the practicality and effectiveness, to a full field test (if applicable).

Procedurally, the prototyping phase of this study (the most important research activity) includes: (i) selecting limited exemplary themes or topics; (ii) designing assessment tasks in a standardised fashion; (iii) anticipating teachers' potential difficulties in the implementation process; (iv) providing detailed procedural specifications; and (v) applying a systematic process of formative evaluation of the products. On the basis of these considerations, a careful design of assessment materials is expected to improve the initial implementation process and ultimately the outcomes.

As was referred to in Chapter 3 (Section 3.5), several studies have been conducted using educational design research as an intervention approach, such as those done by Mafumiko (2006), McKenney (2001), Motswiri (2004), Ottevanger (2001) and Teclé (2006). For example, Mafumiko (2006) examined how micro-scale experimentation can serve as a catalyst for improving the chemistry curriculum in secondary schools in Tanzania. McKenney (2001) explored the possibilities of computer-based support for Science education materials developers in Africa. Motswiri (2004) investigated how to support chemistry teachers in implementing formative assessment of investigative practical work in Botswana. Ottevanger (2001) investigated teacher support materials as a catalyst for science curriculum implementation in Namibia. Teclé (2006) explored the potential of a professional development scenario for supporting biology teachers in Eritrea.

The research model by Mafumiko (2006) (Figure 4.2) is used to inform the research model for the Intervention Study of this study due to its similarities with this study. It is a design study aimed at improving a science curriculum in secondary schools in a similar context to that of Mozambique. Mafumiko's model shows a development process of teacher support materials and student worksheets in chemistry in four different versions following a subsequent design, formative evaluation, and revision steps of prototypes.



**Figure 4.2: The original model by Mafumiko (2006)**



While the first version was developed by the author following design guidelines and specifications of exemplary lesson materials, the subsequent versions (1 to 4) were designed and formatively evaluated by experts and users. The quality of the prototypes was sought through subsequent analysis of the validity, and practicality and effectiveness of the materials. Due to the shortage of time, a full trial of the prototype Version 3 with teachers and students in the classroom did not take place and the intervention was restricted to appraisal by university students and experts. This led to the situation where only the expected practicality and the expected effectiveness of the third version of the prototypes were demonstrated. More evaluation is needed to demonstrate the actual practicality and the actual effectiveness of the intervention.

In the following subsection the educational design research in the context of the Physics assessment materials of this study is further discussed.

#### *4.3.2.2 Design of the Intervention Study*

The study on investigating and improving assessment practices in Physics in secondary schools in Mozambique is characterised by a mixture of a survey and educational design research approaches. The exploratory character of the Baseline Survey previously undertaken and the cyclical nature (design, evaluation and revision) of the Intervention

Study are important means of establishing evidence of a good quality within the limitations of this study. In general, the intervention for this study, that is PAM materials, was developed following an educational design research approach (Figure 4.3).

Baseline Survey	<i>Intervention Study</i>
Preliminary research	<i>Prototyping</i>
	 Version 1 Version 2 Version 3 Version 4
Literature review on Physics classroom assessment and Mozambican education policies. Survey on assessment practices.	Design of the prototypes and classroom tryouts (more emphasis on validity and expected practicality with gradual shift to effectiveness).

- Notes:*
1. Flowchart auto shapes indicate the findings from literature review, Baseline Survey and context analysis.
  2. Block curved arrows indicate cyclical character of educational design research approach.
  3. Increasing gray area means gradual up-scaling of the study.

**Figure 4.3: General research design of the study**

As already reported in Section 4.3 and in Chapter 3, the preliminary research phase, aimed at providing information about assessment strategies used for Science subjects, such as alternative, authentic, formative, performance assessments, and the role of both teachers and students in assessment, was undertaken. The policies of the Mozambican education system, particularly related to classroom assessment, were also reviewed including the developments around the curriculum review for secondary education (see also Chapter 2). Based upon the findings of this Baseline Survey, informed decisions were taken regarding the topics to be investigated, the type of assessment practice to be used as an example for improvement, the teaching, learning approach to be followed, and the assessment strategy to be adopted during the Intervention Study. The findings also provided orientation on the formulation of design specifications that generated the methodological direction for the design of the prototypes and its tryout in the classroom.

After the preliminary phase, the intervention study consisted mainly of the *prototyping stage* where the prototyping process is presented in terms of a series of subsequent design and formative evaluation and revision steps of versions of prototypes. Four versions of prototypes were developed before the final product was constructed. Validity, expected practicality, and expected effectiveness of the draft prototypes were the focus of this stage with the aim of acquiring clear empirical evidence of the performance of both teachers and students during the classroom tryout. Three experts, four teachers, and three university students (also mathematics or Science teacher) appraised the first version. The second version was produced on the basis of the revision made from the first version and was used in a classroom tryout by two teachers and their 62 students. The third version was developed following suggestions from users (teachers and students) and was appraised by two experts in an interactive discussion. As was referred to earlier, the practicality and effectiveness of the third version of the prototypes were only ‘expected’ because this version was only appraised by university students and experts and not via empirical testing. This phase resulted in the fourth and final version of the materials. The analysis of the expected effectiveness of this version was done through an evaluation workshop with university students and teachers including final suggestions from experts. Suggestions on the possible incorporation of the material into the new curriculum under review and consequent possible use by teacher training institutions of the PAM materials were given.

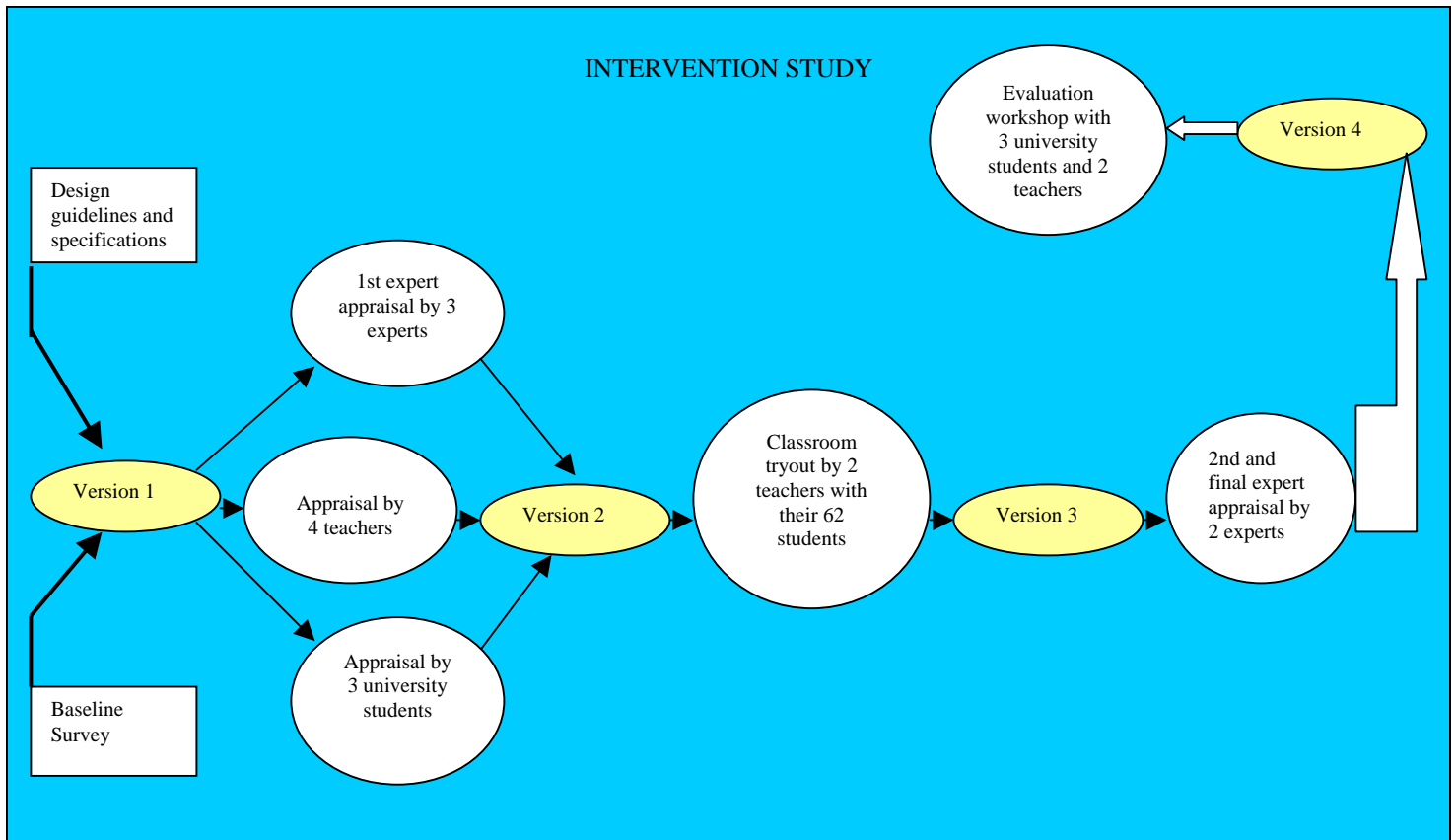
During the development of the four versions of the PAM prototypes, the quality was verified and increased in terms of their validity, expected practicality and expected effectiveness (Mafumiko, 2006; Nieveen, 1999; Ottevanger, 2001; van den Akker, 1999).

1. Validity refers to the internal consistency between the materials and the state-of-the-art knowledge (content validity) and to the fact that the various components of the intervention are consistently linked to each other (construct validity). The first phase of formative evaluation occurred with Version 1 of the prototype through appraisal by experts, university students, and schoolteachers and focused on improving more the validity and less the practicality of the prototype.
2. Practicality refers to the usability of the materials by teachers and students (including experts) in ways that are compatible with the developer’s intention. In

other words, it means performing the design and tryouts of the activities in the conditions put in place in the learning environment, which are prescribed by the current Grade 12 Physics curriculum, and under the schedules of the Physics teachers. The second phase of formative evaluation took place only with Version 2 by classroom tryout by teachers and students and the emphasis was more on expected practicality of the material with little reference to its effectiveness.

3. Effectiveness refers to the extent to which all users (particularly teachers and students) appreciate the experiences and outcomes of the intervention and the learning task. In general, it reveals the implications of the intervention for both teachers and students in light of the acquired theoretical innovations. The effectiveness of the material was verified in Version 3 of the prototype through appraisal by three experts and constituted the third and last phase of formative evaluation. At the end, Version 4 had improved aspects of validity, expected practicality, and expected effectiveness as intended by the intervention.

Figure 4.4 depicts the research model of the Intervention Study, which includes the preliminary and the prototyping phases.



**Figure 4.4: Research model of the Intervention Study**

As was referred to earlier in this section, the model by Mafumiko (2006) was used to inform the research model for the Intervention Study. Three elements in Mafumiko’s original model have been adapted. The first one refers to the use of findings from the Baseline Survey to inform the design process of Version 1. The second element refers to the increased number of appraisers of the Version 1 of the prototypes. The third element is linked to the final appraisal of the prototypes where, due to the shortage of time, it was not possible to try out the final version in the classroom with users to verify its actual practicality and effectiveness.

In summary, the research design for the development of Physics assessment materials for the study on investigating and improving assessment practices in Physics in secondary schools in Mozambique proceeded in the two phases described below.



1. The preliminary research phase consists of a review of the literature about assessment, an analysis of Mozambican education policies, and a Baseline Survey aimed at identifying assessment practices used by teachers in schools.
2. The prototyping phase consists of the development of a number of Physics assessment prototypes to be used by teachers in schools as a way of improving their assessment practices. This phase included:
  - a. the development of the prototypes in cyclical series of design and formative evaluation of the different versions of the prototypes using quality criteria of validity, expected practicality and expected effectiveness of the material in the various prototyping stages; and
  - b. systematic reflection and documentation consisting of analysis of the expected effectiveness of the prototypes and of the sustainability of the study findings.

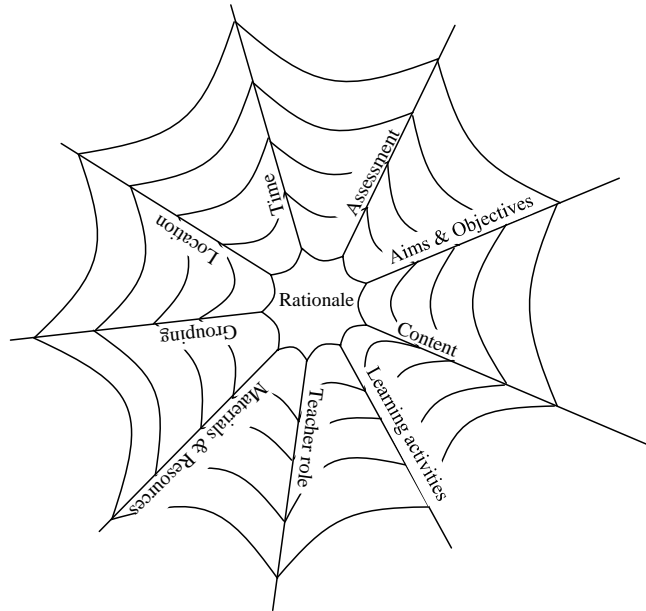
The findings of the preliminary research are discussed in Chapters 2 (Context of the study), 3 (Literature review and conceptualisation of the study) and 5 (Assessment Practices of Mozambican Physics Teachers). Subsections 4.3.2.3 and 4.3.2.4 present the design guidelines and the research procedures for the intervention study. The instrument development for the various formative evaluations and the findings of the Intervention Study are discussed in Chapter 6 (Improving Teacher Assessment Practices in Physics in Mozambique).

#### ***4.3.2.3 Guidelines for design of the intervention***

Before elaborating on the guidelines for designing the intervention on assessment strategies, it is necessary to discuss how the intervention that is being studied can be looked at from a curriculum perspective. The rationale for presenting this curriculum perspective lies in the fact that assessment is always a component of a curriculum and the intervention will necessarily consist of lessons within which assessment will be the focus of interest. This argument is supported by international literature according to which assessment, instruction and curriculum go hand-in-hand and any attempt to predict a

future direction for assessment must consider the factors that can influence curriculum changes (NRC, 2001; Popham, 2002; van den Akker, 2003). Sadler (1989), quoted by the NRC (2001), provides a conceptual framework that places classroom assessment in the context of curriculum and instruction. Within this framework, three elements are required for formative assessment to promote learning, namely: a clear view of the learning goals derived from the curriculum; information about the present state of the student derived from assessment; and action – taken through instruction – in order to close the gap. Popham (2002) refers to the idea that teachers need to know how to create their classroom assessment devices, interpret, and use statewide test results, and then plan instruction based on instructionally informed assessments. This author also supports the view that teachers must be concerned with their instruction and with the fact that their assessments address appropriate content. Van den Akker (2003) argues that the various curriculum components are a powerful tool in understanding the planning of student learning and the development of accompanying learning materials. He describes, in what he calls a vulnerable curriculum spider web, ten curriculum components to consider in curriculum design and implementation and points to the fact that the crucial challenge for curriculum improvement is to establish balance and consistency between the various curriculum components (see Figure 4.5). With the term “spider web” the author’s intention is to illustrate an existing similarity between a spider web and a chain: the spider web is as strong as its weakest part; similarly, a chain is as strong as its weakest link. In fact, focusing on assessment means that the intervention is focusing on one of these components of the curriculum and any effort to improve assessment should be in a balanced and sustainable manner, taking into account all the components of curriculum.

One relevant lesson that can be learnt from these arguments is that, before putting more emphasis on assessment materials, one should focus on the quality of the lesson materials. For the teachers to be able to conduct effective assessment strategies they need support on preparing good lessons and therefore they need materials of good quality.



(Source: van den Akker, 2003)

**Figure 4.5: The vulnerable curriculum spider web**

Each of the ten components of the spider web addresses a specific question about the planning of student learning. Table 4.4 shows the curriculum components and their corresponding focus questions.

**Table 4.4: Curriculum components**

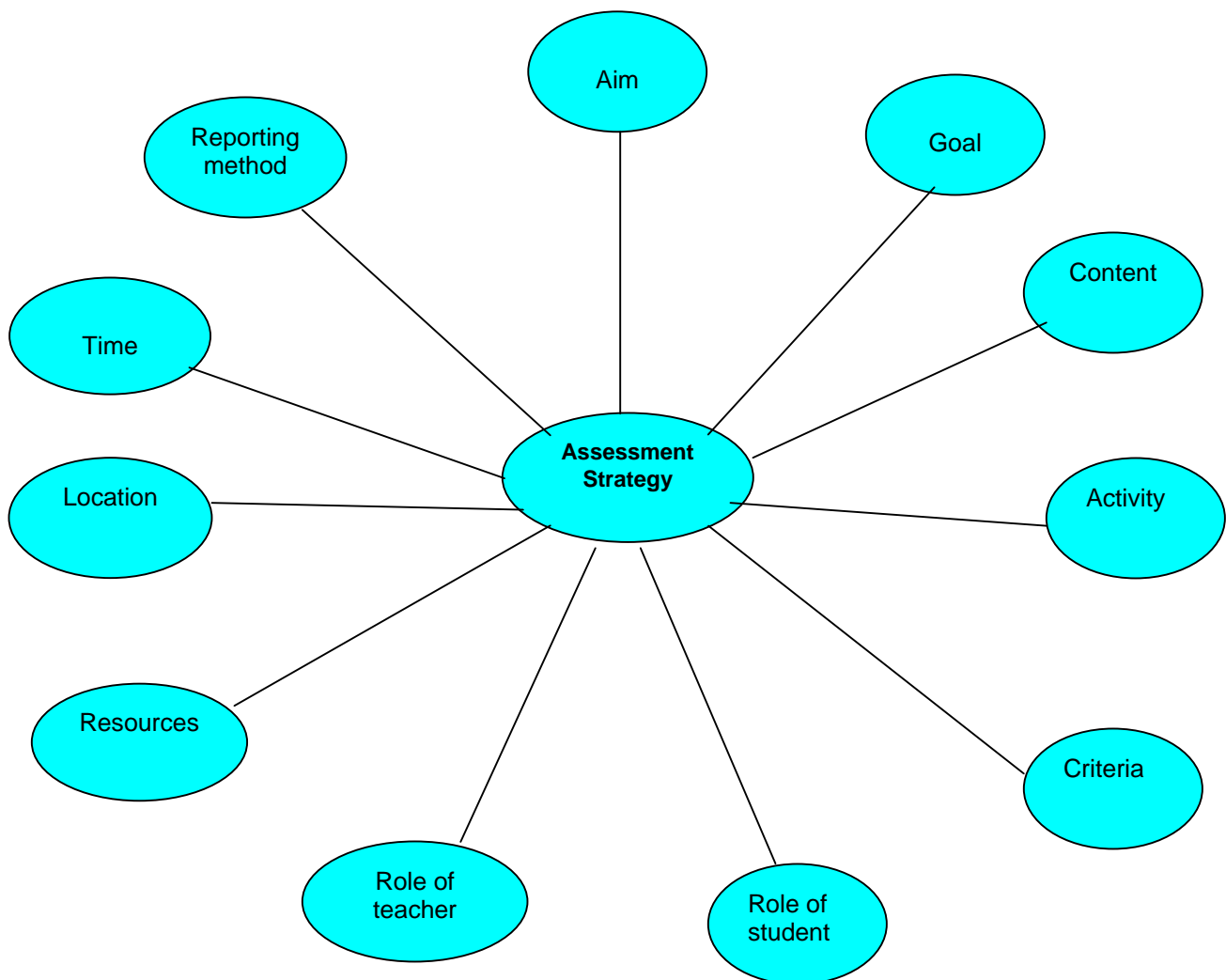
Curriculum component	Focus question
Rationale	Why are the students learning?
Aims & Objectives	Toward which goals are they learning?
Content	What are they learning?
Learning activities	How are they learning?
Teacher role	How is the teacher facilitating learning?
Materials & Resources	With what are they learning?
Grouping	With whom are they learning?
Location	Where are they learning?
Time	When are they learning?
Assessment	How far has learning progressed?

(Source: van den Akker, 2003:4)

The van den Akker curriculum spider web (Figure 4.5) shows the dynamic interactions of various components of a curriculum with the rationale at the centre of the spider. It is used in this study to describe the student-centred approach as the focal point to which the other

nine components are linked. The pertinent question under ‘rationale’ is ‘why are the students learning?’ and this means that the answer to this question has implications for the teaching and learning methods followed, as well as the materials and assessment strategies used. Therefore, the ‘rationale’ is the central mission in the learning process.

Similarly, since assessment practices are the focus of this study, for them to be successful, one may look at the various components of which each assessment can be composed, and thus, the assessment strategy is the central aspect of the assessment process. These components are visualised in the assessment wheel shown in Figure 4.6 (adapted from Howie, 2006).



(Source: Howie, 2006)

**Figure 4.6: Assessment components**

The wheel illustrates that, for assessment strategy to lead to effective learning, several aspects of classroom context must be taken into account and each one must support the other. These aspects are indicated in the wheel in an order (clockwise) which reflects the complete cycle of learning and assessment events advocated by Harlen (2006) and supported by van den Akker (2003). Students must understand what they are supposed to be learning and what is to be assessed and then have a confidence that teachers know these. This understanding includes clarity in aims, goal, content, activities, (assessment) criteria, as well as students' roles and those of the teachers prior, during and after the assessment task. The assessment strategy must also include important planning elements like materials (resources) with which students are supposed to be assessed, where (location) the assessment task will be taking place, and when (time). In the end, the way all these aspects are communicated to students (reporting method) is crucial for effective assessment of learning to take place. Having the assessment strategy at the centre of the wheel implies that the manner in which progress of the student learning can be assessed depends on all these various assessment components. Any changes in the assessment strategy will have to consider changes between its various components as this could affect the student learning.

The quality criteria of the PAM materials discussed in subsection 4.3.2.2 are inherently linked to the van den Akker's typology of curriculum implementation and Howie's interconnections of the assessment component. All these assessment components embody how the curriculum evolves in all its typologies (intended, implemented and achieved) and show the importance of linking assessment (assessing student learning), instruction (what is being taught and how), and curriculum (what should be taught). The validity aspect focuses on the intended curriculum, the practicality aspect focuses on the implemented curriculum, and effectiveness of the materials refers to the achieved curriculum.

In order to reduce the number of assessment components during the classroom tryouts, some adaptations and combinations of the components presented in this wheel have been made. These combinations also ensured that the assessment materials are user friendly for both teachers and students. Thus, in the assessment components of aim, goal, and location

some elements of why the teacher is assessing and in which context the assessment takes place, were added, which resulted in a new component named Rationale and Setting, while content included the student performance expectations and constituted another component called Content and Performance Expectations. Activity, roles of both teachers and students, and time were all embedded in one assessment component named Method, and criteria and reporting method were put into the component of Assessment. The component of resources stood alone as such and was renamed Materials and Resources to imply that these resources include not only books and pencils but also laboratory equipment. This resulted in five assessment components, which are listed below.

1. Rationale and setting: Why is the teacher assessing, toward which goals, and in which context is the assessment component being applied?
2. Content and performance expectations: What content, and on which intended learning outcomes is the assessment focused?
3. Method: (i) What are the activities of the students? (ii) What are the activities of the teacher? (iii) With whom are the students doing the assessment? (iv) At what time in the teaching and learning process is the assessment best applied?
4. Materials and resources: With what materials and resources are the students being assessed?
5. Assessment: How is the quality of the students' final product or task being judged?

Having presented and discussed the arguments in favour of employing certain assessment components in the Intervention Study, the following discussion is about guidelines for designing teacher support assessment materials. Howie (2006) cites Gronlund (1998:18) arguing that during the process of designing assessment materials leading to effective assessment one needs (i) to have clarity on all intended learning outcomes, (ii) a variety of relevant assessment procedures, (iii) fair procedures for all students, (iv) specified criteria for judging students' successful performance, (v) feedback to students that emphasise strengths and weaknesses, and (vi) a comprehensive system of grading and reporting.

Based on these arguments, and as also discussed in Chapter 3 (Section 3.5), it is advisable to design exemplary teacher support assessment materials that focus on four support

levels, namely subject knowledge, lesson preparation, teaching methodology, and assessment and feedback. In the context of this study these support levels are described below.

*1. Subject knowledge*

This level of support is specific to the topics under investigation. It includes making connections with other related topics and a contextualisation in relation to students' prior knowledge focusing on what may support or hinder the students' understanding of the studied topics. It is important that teachers make sure that all these aspects are dealt with before students start to perform the assessment task.

*2. Lesson preparation*

This includes advice to the teacher on the background of the problem that students are expected to solve. It also includes procedural specifications on investigative experimental work, the type of questions to ask while guiding students in the assessment task, and the necessary resources or equipment that students would need.

*3. Teaching methodology*

It includes advice on how to guide students in a student-centred approach for demonstration experiments. This refers to the roles of both the students and the teacher, which includes monitoring how students acquire content knowledge and practical skills.

*4. Assessment and feedback*

This support provides guidance on how to assess the products or characteristics of the students' activities and how to use the results as formative assessment feedback for future planning.

Because Physics support assessment materials are meant to change the teachers' routine and practice by turning their assessment (or even teaching) activities into a more investigative approach, these materials need to be:

- based on the objectives of Curriculum Plan for Secondary Education for Physics, Grade 12;

- developed from materials teachers already use;
- made to reflect clearly stated learning outcomes identified in the core curriculum that students are expected to study;
- designed with adequate support for teaching, and assessment strategies such as lesson preparation, subject knowledge, teaching methodology, and assessment and feedback;
- made to engage students, support curriculum implementation and instruction, improve student teaming, and to report individual student progress; and
- made to help teachers adopt a student-centred approach that includes investigative work and formative assessment.

Turning teacher assessment activities into an investigative approach in the context of this study means having teachers using not only already designed assessment materials but also participate in developing, trying out, evaluating and using their own assessment materials.

Taking into account the assessment components discussed earlier, the following is a discussion of the design specifications for the four support levels as applied to the context of designing the Physics assessment materials.

### *1. Subject knowledge*

Two Physics concepts were the focuses of the prototypes namely force and *inertia* with the aim of helping teachers to assess their students formatively. The two topics were chosen for the following reasons. Firstly, the topics have been identified by the literature as sources of various student alternative conceptions or misinterpretations in many areas of physical Science. Many articles have discussed a number of student alternative conceptions about *force* as related to motion (Champagne et al., 1980; Clement, 1982; Dekkers, 1997; Gunstone, 1987; Thijs, 1987). Dekkers, for instance, lists 19 generalised student statements about situations involving force and the interpretations of their own beliefs expressed in those statements. Due to the fact that in many instances force implies an alteration of state of rest or motion of an object, these student alternative conceptions



constitute difficulties to the understanding of the concept of ‘inertia’. Secondly, within the Grade 11 and 12 Physics Syllabus, these two topics have been given a great deal of attention. They are extensively taught in both grades and, added to the concept of energy, they appear to be the most difficult for students.

The teaching (and assessment) strategy proposed by this study to examine the students’ understanding of the two concepts is Prediction-Observation-Explanation (POE) suggested by White and Gunstone (1992). This strategy requires students to carry out three different tasks. Firstly, they must predict the outcome of some event, and must justify their prediction. Secondly, they must see or perform a demonstration of the event and must describe what they see. Finally, they must reconcile any imbalance or conflict between what they predicted and what they have actually observed. Details entailing each of these tasks are discussed under teaching methodology’s support level.

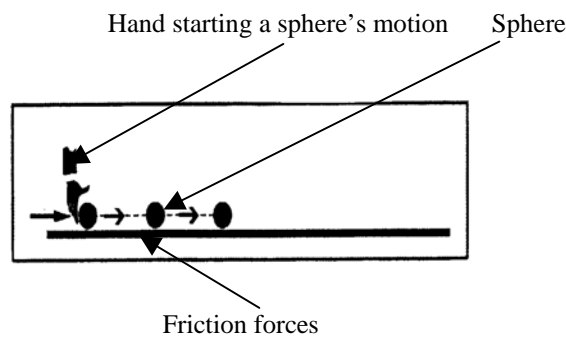
Having indicated the topics used as examples for the demonstration experiments, the rationale of choosing them, and the proposed teaching strategy, the following subsections discuss how these two topics can be introduced (following the proposed strategy) and presents some activities as examples.

#### a) Introduction to the concepts of force and inertia

At the beginning of this subsection the introduction of the force concept is discussed. There are various ways of introducing this concept. Some literature recommends starting the teaching of the concept following a cognitive conflict approach whereby the students’ prior knowledge and understanding (including their alternative conceptions) on the related subject matter is probed (Clement, 1993; Dekkers, 1997; Mutimucuo, 1998; White & Gunstone, 1992). Firstly, and following the recommendation, the probing can start by giving students some examples of a variety of forces, such as gravity, normal or supporting forces, forces in collisions, and forces of springs. At this teaching stage, students do not yet have a clear notion of interactions between pairs of objects but this can later be given attention at Newton’s Third Law of motion. Secondly, students can be helped to understand the concept of force in two conditions: at rest and in motion. When

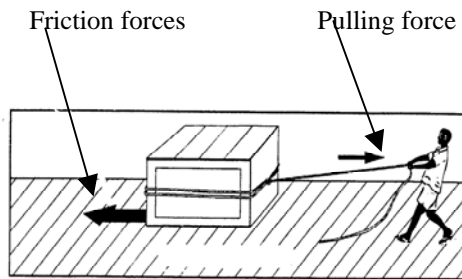
an object is at rest, the sum of all acting forces is zero. For the condition of motion, the sum of all acting forces can be zero (uniform motion) or different from zero. For this study, the activity describes a laboratory demonstration experiment of a moving object with uniform motion. The aim of the experiment is to identify and compare forces acting on moving objects following the POE strategy. This leads to the formulation of Newton’s First Law, according to which, when the sum of all forces is zero, an object at rest or in uniform motion in a straight line will continue in its state, unless it is compelled to change that state by external forces acting upon it.

Two experiments by Galileo can be used to introduce the physical significance of Newton’s First Law of motion. In one experiment, and studying the motion of a sphere moving on a horizontal surface, Galileo observed that, if the sphere were pushed with a given force, it would move through a certain distance before stopping (Figure 4.7).



**Figure 4.7: A sphere set in motion**

When analysing the experiment, it becomes relevant to find out why the sphere comes to rest some time after the pushing force has stopped acting. The reason is that in any moving object (pushed or pulled) there are acting opposing forces, which are impeding the movements. These forces are acting between the moving object and the surface where the object is and they are called ‘friction forces’ (see Figure 4.8).



**Figure 4.8: Forces on a moving object**

Friction forces always act where there is a contact between the moving objects and the surface on which they move and they are forces opposing that movement. In the example of the experiment on Figure 4.7, the friction force acts between the sphere and the horizontal surface and is opposing the movement of the sphere.

In another experiment, Galileo decided to polish the surface on which the sphere was moving to see whether the characteristics of the movement would change. In this case, he realised that the sphere had traversed a bigger distance compared to the previous experiment, before the surface was polished. Then Galileo came to the conclusion that, if it was possible to completely eliminate the force that tends to oppose the movement of the sphere (the friction force), the sphere, after experiencing the action of the initial force (with the push), would continue moving with constant speed in a straight line. From Galileo’s conclusions, Newton formulated his First Law of motion, known also as Law of Inertia, as follows:

*Every object continues in its state of rest, or of uniform motion in a straight line, unless it is compelled to change that state by external resultant forces exerted upon it.*

The key word in this formulation of Newton’s First Law is ‘continues’: an object continues to do whatever it happens to be doing (rest or uniform motion in a straight line) unless an external force is impressed upon it. If it is at rest, it will continue in a state of

rest. If it is moving, it continues to move without turning or changing its speed. Objects at rest tend to stay at rest – objects moving tend to continue moving. This is the physical significance of Newton’s First Law of motion and this tendency of objects to resist changes in motion is called ‘inertia’.

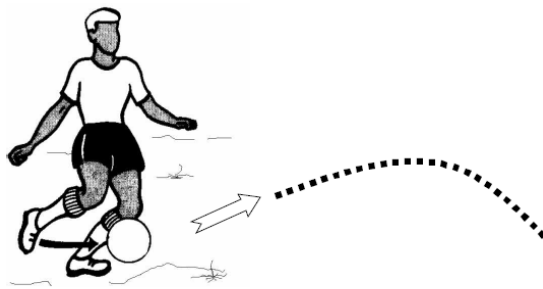
Having discussed the introduction of the force concept, the following subsection discusses how inertia can be introduced. Drawing from the previous discussion about force and once Newton’s First Law is clearly understood, the teacher can easily introduce the concept of inertia. The tendency of objects to remain in its state of rest or of uniform motion in a straight line is called *inertia*. The inertia of objects is observable when (i) an object at rest is suddenly set in motion or (ii) when an object animated with uniform motion in a straight line has the value of its speed or its direction changed. This change is caused by an external influence, which means that the resultant of all forces acting on the object is different from zero. These resulting forces different from zero will cause acceleration. As a conclusion, the inertia of an object will be observable only if the object is accelerated.

Newton’s First Law is another way of showing that all matter (objects) has a built-in opposition to being moved if it is at rest or, if it is moving, to having its motion changed. This property of matter is called inertia. The effect of inertia is evident, for instance, on the occupants of a car which stops suddenly. The occupants will be lurched forward in an attempt to continue moving. The larger the mass of an object, the greater is its inertia; that is, the more difficult it is to move, when at rest, and to stop, when in motion. So, the mass of an object is the measure of its inertia.

Some activities on both force and inertia can be suggested for students to carry out following the POE strategy. Below is an example on how to identify and compare forces on moving objects.

b) An introductory activity using the POE strategy (e.g., identification and comparison of forces acting on moving objects)

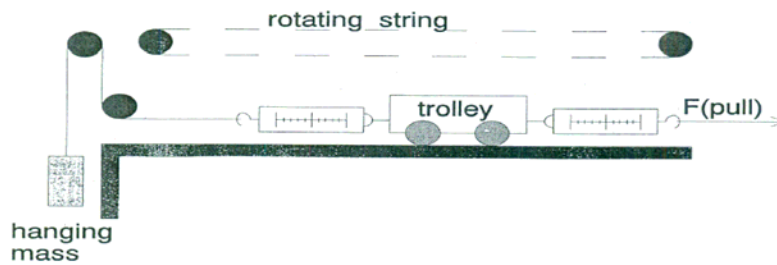
A prerequisite for the demonstration experiment is that students should previously have used spring balances to measure forces, and that uniform motion from the kinematical perspective has been discussed. As a first experiment, the POE strategy recommends that the lesson can start by asking students to name all forces acting on a soccer ball kicked into the air (prediction). After this exercise they can be asked to kick the ball (as shown in Figure 4.9), observe its behaviour through its entire trajectory, and name all the forces acting on the ball (observation).



**Figure 4.9: A soccer ball kicked into the air**

In the end, they must be asked to compare what they have predicted and what they actually observed during the experiment (explanation).

In a second experiment the students can be shown the set-up depicted in Figure 4.10. A trolley is placed on a smooth runway, with spring balances attached to front and back. At the back, a hanging mass is attached to the balance by means of string and pulleys. At the front, the trolley is pulled forward by hand ( $F_{\text{pull}}$ ).



**Figure 4.10: Identification and comparison of forces acting on moving objects**

*Predict:* Using the POE strategy, students are asked to predict, individually, how the forward and backward force will compare, if the trolley is pulled forward at constant speed. They are also asked how these forces will compare, if subsequently a bigger constant speed is chosen.

*Observe:* After the predictions, the experiment is carried out, preferably with the students in small groups. The hanging mass should be big enough for the friction to be negligible. This will ensure that the same forces forward and backward are found at all constant speeds ranging from 0 to 2 m/s (meters per second). A constant speed is obtained pulling the trolley along with a knot in a string revolving above the set-up, which is propelled by an electrical motor from a cassette player. Using various gears, different constant speeds can be obtained.

*Explain:* The result of the observation is compared with what is found in the textbooks about the behaviour of the forces acting on moving objects and it can therefore be named Newton’s First Law of motion. The result of the experiment can then be discussed in the class: ‘What did the students think would happen?’, ‘Which are the acting forces?’, ‘Why were the backward and forward forces equal?’

It is important to mention that, although these experiments are useful in identifying and comparing forces, they cannot necessarily be used to combat potential students’

alternative conceptions on forces. They only offer students the empirical evidence that their expectations and their alternative conceptions can be deficient. They do not provide explanation about what students already know that can guarantee the shift from their own conceptions to the scientific view.

## *2. Lesson preparation*

Teacher support on the preparation of lessons in which the formative assessment will take place includes advice from the teacher on the characteristics of the lesson and student readiness in terms of the background of the problem that they are expected to solve. General characteristics of the Physics demonstration experiments include content specific knowledge (e.g., description of the intended learning outcomes) and procedural specifications of laboratory work (e.g., materials required for the experiments, the timing of the activities, and suggestions on how to deal with potential problems that may occur). In general, the teacher support for lesson preparation includes two main aspects.

### *a) General description of the lesson*

- Firstly, this involves a description of the main concepts (force and inertia) to be dealt with during the experiments and how they will be formatively assessed. The teacher may start the lesson by asking brief questions to students on what they already know about concepts like mass and speed. A discussion about these concepts may help the teacher to understand and evaluate student predictions and/or their responses during experiments.
- Secondly, there is a description of what constitutes the lesson (for instance, that the students will work in groups of a maximum number of four students each).
- Thirdly, there is a description of the intended learning outcomes based on the Physics Syllabus for Grades 11 and 12 - the aims of the lesson must be clearly formulated by the teacher (emphasising the POE strategy) to help students understand what is expected of them.

### *b) Lesson preparation*

- Explain what is to be done and when (lesson plan and timing).

- Explain the working method: for example, explain that everyone in the classroom must be organised into groups of a maximum number of four students each.
- Anticipate potential difficulties during the experiments associated with the student-centred practice.
- Locate the materials (equipment) required for the experiments. This activity is important to make the teacher aware that the experiments are designed to support student-centred practice using locally available materials.

### *3. Teaching methodology*

As referred to earlier, the teaching methodology used to teach and to investigate student understanding of the two Physics concepts is the *Prediction-Observation-Explanation* (POE) strategy. One of the most powerful contributions of the POE strategy to learning is that it is more direct in revealing students' understanding than the usual style of verbal or paper-and-pencil tests. It focuses on a specific phenomenon of learning. The prediction that is required from students is more likely to imply genuine application of the previous knowledge rather than asking a simple question in the form “explain why...”. Furthermore, the students are more likely to evaluate how their knowledge applies to a real situation, because the experiment is directly shown, than the more general thinking implied by a single question. Another key characteristic of the POE strategy is that it allows students to decide what reasoning they must apply in any given situation, whilst their predictions are based on their everyday experiences and beliefs.

Besides the three main tasks that students are required to carry out (predict, observe, explain), some critical steps are to be taken into account when applying this strategy (White & Gunstone, 1992). The first step is that teachers must ensure that all students understand the nature of the situation about which they are supposed to make a prediction. Teachers should also ensure that all students have the same understanding of the situation before proceeding. This can be done by encouraging students to ask questions about the situation under consideration. The second step refers to the importance of having all students indicate, in writing, both the prediction and the reasons supporting the prediction.



This is important because it allows the students to decide what knowledge is appropriate to apply and then apply it. The way students must indicate their prediction and the supporting reasons can be either in open-ended form, i.e., having students writing their predictions on their own words, or on a previously prepared sheet of paper. Recording the reasons for prediction is crucial for the value of the teaching strategy because it shows the link between the concepts involved in the learning situation. The third step occurs during the actual experimentation. It is important that all students write down their individual observations while the experiment takes place. Very often different students will see different things, and if observations are not written down at the time they are made, some students might change their observations as a result of hearing what others claim to have seen. The fourth and last step refers to the students' reconciliation of any discrepancy between what they predicted and what they actually observed. Normally this is difficult for students, but it is advisable because students' explanations at this stage reveal much about their understanding.

#### *4. Assessment and feedback*

The ultimate objective of this support level is defined in terms of assessment *of* learning and thus, the emphasis is on all functions of assessment namely diagnostic, formative, and summative. The diagnostic and formative functions of assessment are expressed by a number of design guidelines to facilitate experimental work and of elements of feedback provision that the teacher needs to consider. These design guidelines, as taken from the literature (Dekkers, 1997; Garrett & Roberts, 1982; Gunstone, 1991; Tamir, 1991; van den Berg & Giddings, 1992), are listed below.

- Agreement – having stated the problem to be investigated, the teacher and the students must agree on the procedures to be followed, the evaluation of the explanations given during the experimental work, and the conclusions.
- Learning outcomes – the teacher must be tightly prescriptive about the ideas that the students are supposed to acquire and develop.

- Student participation – In practical work, particularly in laboratory demonstrations, the teacher must produce the event to be investigated according to the purpose to be achieved, while the students attempt to interpret it and make sense of it.
- Type of experiment and aims - Teachers must avoid having too many aims of the experiment to be achieved at once. This may lead to none being pursued.
- Critical thinking and reporting – Teachers are to make sure that students develop a critical attitude towards their actions and interpret the activity’s data only in the light of the experimental work pursued and of their own knowledge and experience.

As for providing formative feedback to students, when facilitating demonstration experiments, teachers must consider a number of elements of feedback provision in three main stages of the lesson, namely (i) in lesson preparation, (ii) in the course of the lesson, and (iii) in the end of lesson (Motswiri, 2004). These elements are presented and discussed in Chapter 6 (Section 6.5).

Concerning the summative function of assessment, the use of assessment criteria is crucial to help monitor the performance of students. The criteria to be adopted for assessing student understanding of the inertia concept using laboratory experiment must be such that they provide information about how students performed the task at the end of the experiments. Scoring rubrics are used to assess the student responses to the performance task. These rubrics are observable in nature, and there are specific aspects a student should perform to carry out a performance properly.

Two types of scoring criteria are frequently discussed in the literature, namely analytic and holistic (Moskal, 2003). Analytic scoring rubrics divide a performance into separate facets and each facet is evaluated using a separate scale (see, for example, the different performance levels of the rubric on assembling an electric circuit in Table 3.2). Holistic scoring rubrics use a single scale to evaluate the larger process. In order to develop observable scoring criteria for the POE strategy, analytic scoring rubrics were considered

(refer to Appendix P, Table 2) and the guidelines discussed below were taken into account (Moskal, 2003:15).

- *The criteria set forth within a scoring rubric should be clearly aligned with the requirements of the task and the stated goals and objectives.* A list can be compiled that describes how the elements of the task are into the goals and objectives. This list can be extended to include how the criteria set forth in the scoring rubric map into both the elements of the task and the goals and objectives. Criteria that cannot be mapped directly back to both the task and the goals and objectives should not be included in the scoring rubric.
- *The criteria set forth in scoring rubrics should be expressed in terms of observable behaviours or product characteristics.* A teacher cannot evaluate an internal process unless this process is displayed in an external manner. For example, a teacher cannot look into students' heads and see their reasoning process. Instead, it is necessary for students to explain their reasoning in written or oral form and the scoring criteria should be focused upon evaluating the written or oral display of the reasoning process.
- *Scoring rubrics should be written in specific and clear language that the students understand.* One benefit of using scoring rubrics is that they provide students with a clear description of what is expected *before* they complete the assessment activity. If the language employed in a scoring rubric is too complex for the given students, this benefit is lost. In other words, students should be able to understand the scoring criteria.
- *The number of points that are used in the scoring rubric should make sense.* The points that are assigned should clearly reflect the value of the activity. On an analytic scoring rubric, if various facets are weighted differently from other facets of the rubric, there should be a clear reason for these differences.
- *The statement of the criteria should be fair and free from bias.* The phrasing used in the description of the performance criteria should be carefully constructed in a manner that eliminates gender and ethnic stereotypes. Additionally, the criteria

should not give an unfair advantage to a particular subset of students and which is unrelated to the purpose of the task.

#### ***4.3.2.4 Research procedures for the intervention study***

This subsection presents the research procedures for the Intervention Study. It describes the number of activities carried out during the design and development phases of the PAM prototypes and the period in which these activities took place. Details about how appraisers of the materials and the participating schools, teachers, and students were prepared for their roles in the intervention are all presented in different sections in Chapter 6 according to their level of involvement (refer to Sections 6.2 to 6.5).

The Intervention Study consisted of a cyclical development of the PAM prototypes in four versions and was undertaken between April 2006 and February 2007. The first version of the PAM prototype was designed by the researcher based on (i) lessons from Baseline Survey findings and (ii) design guidelines and specifications of exemplary assessment materials described earlier (subsection 4.3.2.3), which were adapted from design specifications used by similar intervention studies in Science education (refer to Chapter 3, Section 3.5). This version was appraised by three experts, three university students, and four secondary school teachers. The design and appraisal of this prototype was carried out from April to June 2006. The design and development of the second version – also by the researcher - was undertaken between July and September 2006 and followed suggestions and recommendations from appraisers of the first version. This version was then tried out in the classroom with two teachers and 62 of their students in October 2006. Then the analysis of the tryout findings, which started during the tryout in October, was finalised earlier in November 2006. The findings led to the design - also by the researcher - of the third version of the prototype. It then followed a final appraisal by two experts between November 2006 and January 2007. Finally, the fourth and final version of the prototype was designed and evaluated in a workshop with three university students and two teachers in February 2007.

#### **4.4 Validity and reliability**

No research study is perfect. However, controlling the possible threats, which might interfere with the interpretation of the cause-effect relationship, is crucial (Coolican, 1999). Understanding the concepts of validity and reliability helps to analyse the possible weaknesses derived from uncontrolled variables particularly in experimental research. The general definition of validity is that it is a demonstration that a particular instrument does indeed measure what it is supposed to measure (Cohen et al., 2000). Reliability is defined by Cohen et al., (2000) as a synonym for consistency and replicability over time, over instruments and over groups of respondents.

For this particular study, three means were considered to establish validity. The first was face validity where for both research phases (baseline and intervention) the validity was checked in all the corresponding data collection strategies, namely questionnaires, interviews schedules, classroom observation schedules, including the evaluation instruments of the classroom tryouts. The validity was checked by inspecting whether the instruments indeed measure what it is supposed to measure in terms of level and breadth.

The second was content validity (Cohen et al., 2000) of the instruments, which was controlled through consultation with colleagues, teachers, experts, and also by relying on the researcher's experience to ensure the representativeness of the researched area.

The third was external validity (Yin, 1994) of the intervention, which deals with the problem of knowing whether the demonstration experiment results can be generalisable to a broader perspective. This is a particularly difficult notion to achieve for a case study like the one reported in this dissertation. Although no statistical generalisation is possible from the sample involved in the study to the population of Mozambican Grade 12 Physics teachers and students, yet this study strives to generalise the findings to the broader theory underlying the design and development of the intervention. Yin (1994) speaks in this context of analytical generalisation. There will have to be more replications of these findings in more classroom tryouts to determine whether the same results may occur.

In terms of reliability the issues of internal and external reliability were addressed. Internal reliability or consistency was verified to ensure that questionnaire respondents answer related items in similar ways. External reliability or stability was verified by cross checking information on assessment practices used by teachers by comparing the information provided in the questionnaires to that of the interviews and including the classroom observations.

#### **4.5 Ethical issues**

The subject of ethics in social research is potentially a wide-ranging and challenging one. Therefore, it was fitting for this study to address the main issues that may confront the researcher in the field. However, before discussing the issue of confrontation with the field, including the researcher's own integrity and transparency, it is important to state that the ethics requirements, as prescribed by the University of Pretoria, were met. Prior to the research, permission were sought from the Faculty of Education regarding ethical considerations involved in the study. The procedures suggested were approved by the Faculty and permission to undertake the study was granted.

Regarding the issue of field confrontation and researcher's integrity, four aspects deserved consideration for the course of the research.

*a) Debriefing and right to non-participation* (Coolican, 1999) – All participants in this research were informed prior to the research about the full nature and rationale of the study they were to be involved in, and there was an effort to avoid any negative influences. The researcher had to emphasise the voluntary nature of the participation, as well as the right of the participants to withdraw at any time should the discomfort be greater than anticipated.

*b) Confidentiality and privacy* - The researcher guaranteed anonymity or requested permission to identify individual participants. For example, when the use of tape recordings appeared to be necessary, permission was sought. Interviewees and teachers observed during classes and those who participated in the tryouts were all asked for their permission to be identified. For the particular aspect of recorded interviews, the

participants were given the right to assume that the records of the interviews would be safeguarded and used as anonymous data only for research purposes.

*c) Intervention* (Coolican, 1999) - Since the design and development of assessment prototypes was an activity that caused alteration to the teachers' normal routine, there was a need to improve some working conditions. For instance, coffee and snacks were arranged for the teachers in such situations where they had to stay in school much longer than their normal time schedule.

*d) The role of the researcher* (Plomp, 2006) - This research was conducted in close collaboration with teachers and students who were actively involved, often as members of the research team. The situation led to problems of finding a balance between the role of the researcher as a designer, an evaluator, and an implementer. Making the research open to scrutiny and critique by educational experts, deserving attention to validity and reliability of data and instruments, and having a good quality of research design appeared to be key measures for this potential conflicting role. For instance, the quality of the design was sought through triangulation (of data and its analysis), empirical testing (of the intervention), and systematic documentation, analysis and reflection (of the design, development, and evaluation of the intervention process).

## 4.6 Conclusion

The research design of the study on the investigation and improvement of assessment practices used by Grade 12 teachers in Physics consists of two main stages. The first stage, the Baseline Survey, involves the identification of assessment practices currently used by the teachers in Mozambique. The survey is based on the context in which Physics teachers are working in schools as well as on the insights of the literature on what is deemed to be good classroom practice. The literature review (as presented in Chapter 3) focuses on the role, current practices, and ways of improving the teacher assessment practices in secondary Science education as surveyed in many educational systems in both developed and developing countries. Expert appraisal and networking with people working in similar fields have also added value to the preliminary research of the overall study. This stage led to the necessary groundwork to the following stage of the research study. The overall findings of this stage are discussed in Chapter 5.

The second stage of the research design is the Intervention Study, which consists of the design and development of prototypes for assessing the performance of Grade 12 Physics students as a way of helping teachers to improve their classroom practice. The prototypes consist of demonstration experiments and a written report and they were designed and developed on the topics of force and inertia, selected from Grade 12 Physics Syllabus for Secondary School in Mozambique. These topics were chosen because of their suitability to apply the POE strategy in inquiring student understanding of Science concepts, particularly for performance assessments (to be discussed in Chapter 5). The development of the prototypes uses a cyclic approach of design and formative evaluation in such a way that successive versions of the material evolve into a final product with empirical evidence of its practicality. The validity, expected practicality and expected effectiveness of the material were verified through appraisal from curriculum, science, and assessment specialists and tryouts by potential users, i.e., teachers and students. The design and appraisal of the subsequent versions of the prototypes, as well as the findings of the classroom tryout, are presented and discussed in detail in Chapter 6.