CHAPTER ONE

Introduction and Overview

> …while volumes of data are extruded about and from schools, teaching continues without the benefits of such data. There is still a philosophy that assumes teachers know how and what data to collect to best enhance learning, and many of these assumptions are based on folk philosophies, poor measurement, and shaky data. We still teach in a manner we did 150 years ago … (Hattie, 2005, p. 11)

Globally a wealth of educational data has been collected on learner performance, but while this has had a major impact on systemic and curriculum reform (Howie & Plomp, 2005) it seems to have had only limited influence on learning and teaching in classrooms (Fullan, 2006; Hattie, 2005; Patton, 1991, 1997; Visscher & Coe, 2002). It is a global phenomenon that even when school data are available, schools often do not use them or are unable to interpret them correctly and consequently, use the data for improvement of quality of education in the schools (Schildkamp & Teddlie, 2008; Wohlstetter, Datnow, & Park, 2008). This can also be observed in the Republic of South Africa (RSA), where despite the available educational data, formal monitoring and use of data to inform planning, teaching and learning in schools is still limited (Department of Education, 2002b, 2006a, 2006b).

This thesis aims to bridge this gap between the availability of learner performance data and their use in informing planning and action in schools. The thesis employed a design research approach to determine design guidelines and principles[2] to facilitate the use of feedback[3] from learner performance monitoring and enhance the use of data at school level. Design research has two purposes:

[2] Van den Akker (1999) describes design principles as the major knowledge gained from design research. Design principles can be either substantive or methodologically orientated and aim to act as guidelines to other designers faced with a similar problem to support their design tasks.

[3] The use of feedback in this case is not seen as only the use of monitoring information, but also any process use of the skills and knowledge gained by participants during the monitoring process. This conceptual use may include changes in how assessment is viewed, different approaches to problem-solving, and greater confidence in interacting with data, etc. The issue of use of feedback is explored in depth under the conceptual framework in Chapter 3.

1. To design, develop or optimise a product or intervention for a complex real-world problem
2. To develop design guidelines to contribute to the body of knowledge on a phenomenon and support other designers faced with similar challenges in their own context

   (De Villiers, 2005; Plomp, 2009; Van den Akker, 1999)

For the purpose of this study, an existing feedback system known as the South African Monitoring system for Primary schools (SAMP) was optimised (see Section 1.2). The process of developing and improving the feedback system was used to examine ways of facilitating the use of the feedback in participating schools. Design guidelines were produced to facilitate use of learner performance feedback systems. The study examined not only data-use, but also implicit changes in aspects such as problem-solving, planning, data-literacy (see Section 1.1 for definition) and attitudes towards evidence-based practice in the schools.

This chapter first provides the definitions of the terms as used in this study, given the plethora of terms in this field (Section 1.1). Thereafter, the SAMP project that is the focus of this study is discussed (Section 1.2). This is followed by an illustration of the study's importance and value, with the problem presented in context and the rationale given (Section 1.3). The general research questions are introduced in Section 1.4 to be operationalised in Chapters 3-8, based on the literature review, conceptual framework and research procedures. Next, the research methods and design are briefly discussed to frame the empirical component of the inquiry (Section 1.5). This is followed by a discussion of the researcher's role and positioning in this research (Section 1.6). The chapter concludes with an outline of the remaining chapters that comprise this thesis (see Section1.7).

1.1 Definition of Terms

The literature for this study employs various terms such as monitoring, evaluation, assessment, quality assurance, feedback, data, data-literacy, feedback systems and feedback facilitators, all of which refer to measurement

and the use of information. The definitions of these terms are constantly being debated, but working definitions are presented here to provide a common understanding of use in this thesis.

Traditionally, ***evaluation*** is seen as systematic gathering of information that results in judgement (Beeby in Husėn & Tuijnman, 1994, p. 1; Scheerens, Glas, & Thomas, 2003). The Joint Committee on Standards for Educational Evaluation (1994) based their definition of evaluation on the work of Scriven (1967, p. 39) who states that evaluation is "[t]he systematic investigation of worth or merit of an object". For the purpose of this study, these two sources are drawn on to define evaluation as the systematic gathering of data to investigate the worth or merit of the feedback system.

Monitoring involves the tracking of change over time that demands systematic and regular procedures for the collection of data at multiple points in time for decision-making (Husėn & Tuijnman, 1994; Nuttal, 1994). Scheerens et al. (2003) see monitoring as a type of evaluation that calls for ongoing information gathering focused on description rather than valuing and judgement. For the purpose of this study, monitoring is defined as a type of evaluation that requires a systematic collection of data at multiple points in time for the purpose of decision-making. Monitoring focuses on providing information that participants can use to make decisions and apply value and judgement. The aim of these decisions is to improve or enhance education.

Quality assurance is described by Scheerens et al. (2003) as the purpose of monitoring in education. In this thesis, it relates to monitoring and evaluating the quality of teaching, learning, planning and action taken in schools to improve learning.

Assessment is a measurement activity that usually refers to the measurement of learner achievement in the context of education (Joint Committee on Standards for Educational Evaluation, 1994). Hattie and Jaeger (1998, p. 116) note that the difficulty with assessment is that knowledge acquisition is mostly assessed while "…other functions of learning such as deep understanding, efficient intuitive use, acquiring multiple flexible

strategies, adaptive action control, and achievement motivation" are ignored. Assessments are often employed for the purposes of monitoring and evaluation. In this study learners' phonics, reading and mathematics skills are assessed using the South African Monitoring system for Primary schools (SAMP). Therefore, assessment refers here to the baseline and follow-up measurements conducted with learners in SAMP. This forms the basis of the data used in the feedback system.

In the context of education, ***feedback*** relates to returning data gathered through monitoring or evaluation to schools, learners or parents. Hattie (2005) states that feedback can be complex, in particular when considering how best to present data in a comprehensible and accessible form to schools. The aim of attending to data presentation and feedback mechanisms is to facilitate use of the feedback to lead to improvement in schools.

Data[4] consist of discrete, objective evidence collected by qualitative or quantitative methods. Data could be drawn from observation, anecdote, opinions, or figures (such as averages). To qualify as data, systematic inquiry, organisation and analysis that incorporate various views is required (Earl & Katz, 2006). Interpreted data are known as 'information' and have meaning, relevance and purpose (Schildkamp & Kuiper, 2009). In this study, learner performance data were generated through the SAMP assessments. Evaluation data from expert evaluators, teachers, heads of department (HoDs) and principals were also generated through questionnaires, observations, interviews, evaluation reports as well as the Delphi technique.

[4] Data is a Latin plural (singular datum), but is generally defined as uncountable and therefore may take the singular form of the verb.

Data-literacy in this thesis refers to a school's ability to interact actively with data, interpret and apply them. According to Earl and Katz (2006, pp. 19-20) a data-literate school leader is able to:

- think about the purpose for which the data were generated
- recognise sound and unsound data
- be knowledgeable about statistical and measuring concepts
- recognise that there are several types of data
- make interpretation paramount
- pay attention to how data are reported and for which audience

Feedback systems are information systems that focus on generating and providing data to users. In the educational context a specific brand of feedback system can be identified, namely School Performance Feedback Systems (SPFSs), which are "external to schools that provide them with confidential information on their performance and functioning as a basis for school self-evaluation" (Visscher & Coe, 2002, p. xi). The SAMP project on which this study is based is classified as a School Performance Feedback System (SPFS).

In this thesis, a ***feedback facilitator*** is conceptualised as the person responsible for coordinating the monitoring process, compiling reports, and providing feedback to participants, as well as supporting interpretation and encouraging action based on the monitoring and feedback process. In this case, I, as researcher also fulfil the role of monitoring facilitator, but this role may eventually shift to the district office of the Department of Education (DoE) or even individual school coordinators, as the SAMP system develops and becomes more widespread. Sections 1.6 and 4.3.1 provide further discussion of the dual role of facilitator and researcher in this study.

1.2 The SAMP Project

Data for educational improvement purposes can be generated in multiple ways and be managed either externally or internally. Data can also be aggregated and related to different levels of education, including the systemic,

school and classroom levels. Different types of data such as those on learner performance, teacher performance and management functioning can inform educational improvement. This research focuses on optimising a feedback system for a school-based monitoring system, SAMP, which is currently facilitated externally by the Centre for Evaluation and Assessment (CEA). The system produces learner performance data that are also aggregated to school level. The data from SAMP are employed to inform individual learner intervention, classroom practice, and school level planning and action.

SAMP focuses on the entry level to primary education known as Grade 1 (usually five to seven years of age). This forms part of the Foundation Phase of Education, designed to establish basic literacy, numeracy and phonics skills so that learners can progress to learn more independently in the later phases of schooling. V. Greaney (personal communication with SJ Howie, April 18, 2006) stresses the importance of the establishment of foundational skills, since learners who are not able to read at 12 years of age are unlikely to learn to read later in life. Poor foundational skills have also been noted as a large contributory factor to poor learner performance in secondary education by the DoE (Department of Education, 2006b, 2006d)

SAMP produces reliable and valid data for the South African context across the three languages in which it is currently employed, namely English, Afrikaans and Sepedi (for a fuller discussion of the quality of the instrument and current functioning, see Chapter 2). SAMP therefore provided relevant and appropriate data to use in the feedback system and offered the opportunity to study the use of the feedback.

SAMP is a research project originally funded by the National Research Foundation (NRF) to develop monitoring systems at primary and secondary school levels. It was initiated in 2003 in collaboration with the Centre for Evaluation and Monitoring (CEM) at the University of Durham in the United Kingdom (UK). The collaboration aimed to develop a monitoring system suitable for South African education and to develop expertise and resources in the field of learner performance monitoring. The CEM developed a suite of value-added assessments in conjunction with teachers and Local Education

Authorities (LEAs) staff in the UK. The suite employs a combination of objective assessments and teacher ratings to provide valuable information about each learner to schools and teachers (Tymms & Albone, 2002).

The CEA decided to focus on the CEM's assessments for the beginning of primary school (Performance Indicators in Primary Schools or PIPS) and the beginning of secondary school (Middle Years Information System or MidYIS). The entry phases for primary and secondary school were identified as periods when South African schools have limited information about the levels of knowledge and skills of their learners, since learners are often from a large and diverse feeder area (Scherman, Archer, Howie, & Lopez, 2006). The primary school counterpart of the CEM suite in South Africa is known as SAMP, while the South African secondary school component is known as the South African Secondary School Information System (SASSIS).

The PIPS assessment was originally developed in England by the CEM in 1994 (Tymms, Merrell, & Jones, 2004) with the aim of providing not only data on current attainment of learners on curriculum aligned subtests, but also providing predictive data on future mathematics and reading performance (Tymms, Merrell, & Henderson, 2000). The PIPS monitoring system utilises two assessments: a baseline assessment implemented at the beginning of the year and a follow-up assessment administered at the end of the year. This means that both the learners' current level of performance and the difference between the baseline and follow-up performance are reported (for a description of value-added measures, see Section 2.2.1). In 2002, more than a million primary school learners across the globe participated in PIPS (Tymms & Coe, 2003). Figure 1.1 outlines the history of the SAMP project and instruments.

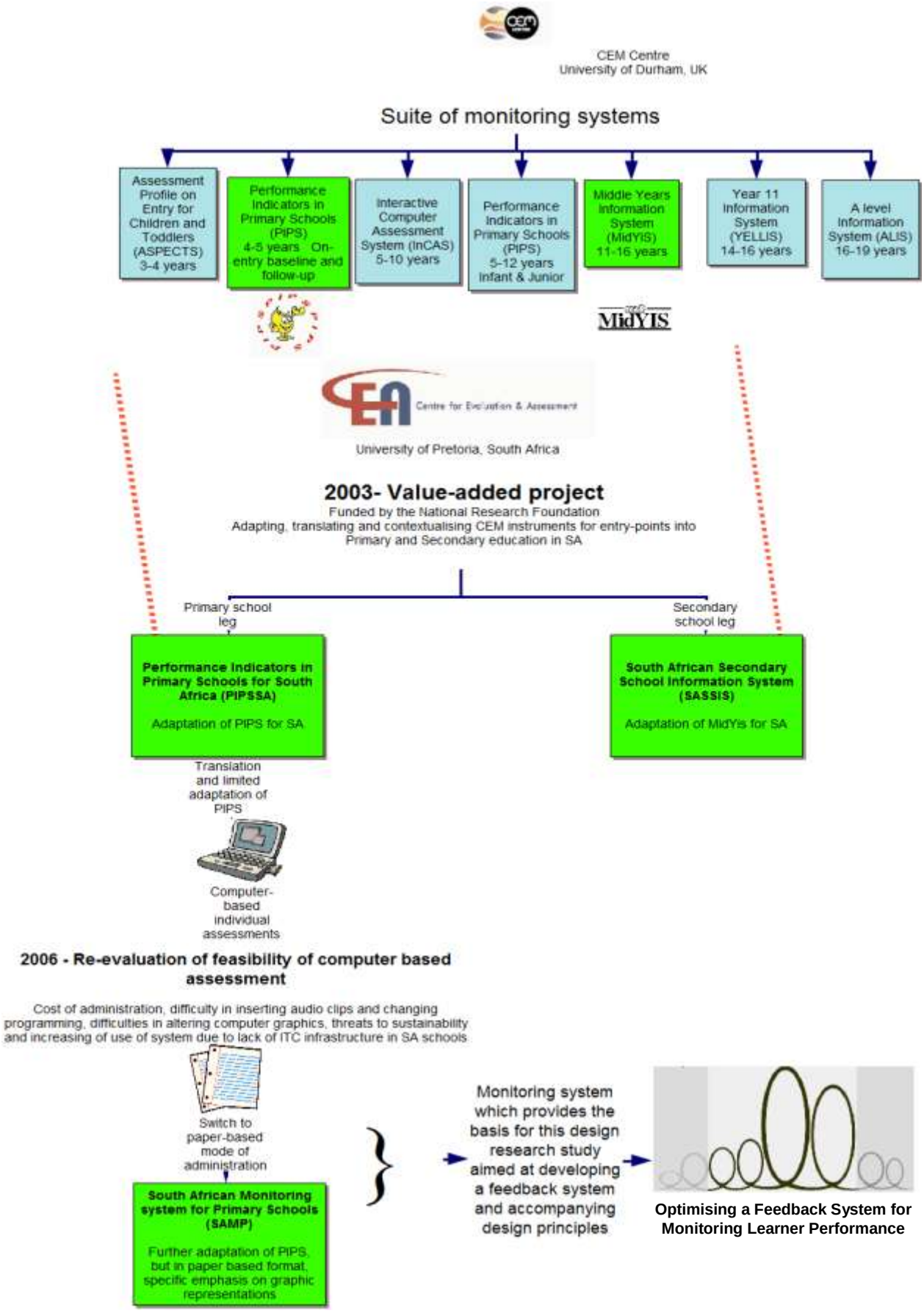

Figure 1.1: History of the SAMP project

The translation and adaptation process of PIPS was initiated in 2003 and originally centred on the computer-based mode of administration of the PIPS instrument, known as Performance Indicators in Primary Schools for South Africa (PIPSSA). During the first three years, the assessment was translated into Afrikaans and Sepedi[5] to determine the feasibility of using the translated PIPS computer-based assessment with the original graphics and instructions in the South African context. The items were also assessed in terms of the national curriculum. The computer-based assessment was recorded in Afrikaans, Sepedi and South African English. These three languages represent the dominant Languages of Learning and Teaching (LOLT) in the Tshwane region where fieldwork took place (H. Julies, DoE, personal communication, July 16, 2010).

The research for this thesis commenced in 2006, the first part of the year being dedicated to consolidation and use of the computer-based assessment data from the previous three years (2003-2005). The project had reached a critical point where the use of the computer-based assessment was not sustainable and limited the expansion or further adaptation of the project. To address this, the next three years of adaptation and development of the PIPS instrument to the South African context focused on paper-based assessment (Archer, 2006a). The project name was also changed to the South African Monitoring system for Primary schools (SAMP) at this stage. The switch to a paper-based assessment immediately increased the feasibility and sustainability of the project in the country's schools as the necessary information and computer technology (ICT) was not yet in place to allow for sustainable computer-based assessment (Department of Education, 2003b; Gauteng Department of Education, 2005). (See Chapter 2 for a full discussion)

The contextualisation process for SAMP was extensive, with various phases of discussion, adaptation and further development of the items incorporating expert appraisal (including Foundation Phase academics, instrument developers, educational and research psychologists and teachers) as well as

[5] A Zulu translation also took place but development was not continued and the SAMP assessment is currently only administered in three languages: English, Afrikaans and Sepedi.

statistical analyses (Archer, 2006a). This process is discussed in full in Chapter 2.

The adaptation process culminated in the SAMP instrument that was employed in 2008 to assess 1,535 learners in English, Sepedi and Afrikaans. The instrument assessed the handwriting, phonics, early reading and early mathematics skills of the Grade 1 learner. Given the South African context, an English Additional Language instrument for non first-language English speaking learners was also developed to assess proficiency in English. The SAMP assessment in its paper-based format consists of a number of subtests, constituting four scales (Early Phonics, Early Reading, Early Mathematics and Handwriting) and the English Additional Language Assessment (as illustrated in Figure 1.2).

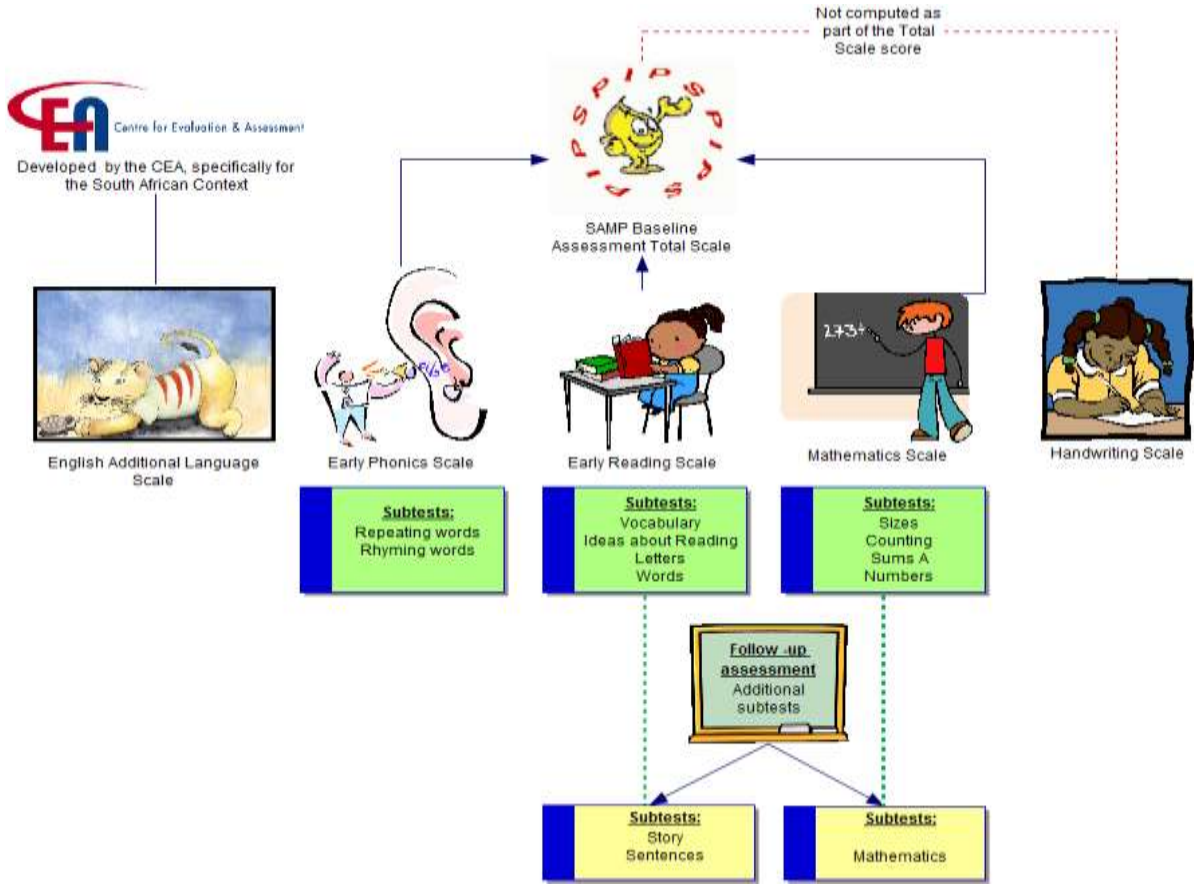

Figure 1.2: SAMP Baseline and follow-up assessment

The baseline assessment consists of 12 subtests[6]; while the follow-up assessment contributes a further three subtests, as indicated in Figure 1.2. The results from the follow-up and baseline assessment are compared and reported on to the schools in terms of gains or losses made.

As a result of the rigorous development and adaptation process (discussed in Chapter 2), the SAMP instruments can be used with confidence as valid and reliable instruments in the South African context. It is therefore appropriate to examine the use of a feedback system based on the data generated by SAMP. In the next section the problem statement and rationale for this study are explored.

1.3 Problem Statement and Rationale

> … [South Africa] is a country with natural wealth and many cultures. It is also notorious for the Apartheid (sic) policies that have left a lasting impression on the education system in the country. Evidence of this [lasting impression] lies in the appalling conditions in many schools across the country, and these conditions exist primarily in previously so-called African, coloured and Indian schools. South Africa, since the first democratic elections in 1994, has embarked on a substantial reform effort in many areas including education. (Howie, 2002, p. 9)

Education is a major concern for the South African government, which had invested 5.8% of the gross domestic product (GDP) in this sector between 1995 and 2003 (National Treasury Republic of South Africa, 2005). In 2006 this investment dropped slightly to 5.4% of GDP, representing 17.6% of the total government expenditure (World Bank, 2008). On average, other upper middle income countries and Sub-Saharan Africa countries spent 4.1% and 4.2% of their GDP respectively on Education in 2006 (World Bank, 2008). The National Treasury of RSA (2005) noted that there had been an enormous growth in enrolment figures for primary and secondary schools (8.1 million in 1985 to 12.0 million in 2004). Population growth and immigration had no doubt contributed to these figures, as the net enrolment rate in primary education dropped from 90% in 1991 to 88% in 2006 (World Bank, 2008). In

[6] The Handwriting and English additional language assessments consist of only one subtest each, which are also used as the scale scores.

2006, South Africa still had nearly 470,000 primary school children aged children who did not attend school (World Bank, 2008).

Despite the significant funding and increase in enrolment for education in South Africa, the quality of education remained a concern (Taylor, Muller, & Vinjevold, 2003). Nowhere was the shortcoming of education provision more apparent than in the low learner performance, especially in subjects such as Reading, Mathematics and Science. This low learner performance was clearly illustrated through South Africa's performance in international studies such as the Trends in Mathematics and Science Study (TIMSS) 2003 (Martin, Mullis, Gonzalez, & Chrostowski, 2004) and the Progress in International Reading Literacy Study (PIRLS) 2006 (Howie, et al., 2008). The concerns about South African education are further highlighted in national studies such the Grade 3 and 6 National Systemic Evaluations (Department of Education, 2002b).

It may be that this poor performance is a legacy of the apartheid education system, however, there is evidence of an international trend wherein increased investment in education is not necessarily associated with improvement in education (Cassassus, 2001; Hayward & Hedge, 2005). Hattie (2005, p. 12) notes that in the United States of America (USA) "…there is not a lot of evidence that the massive increases in state/federal monies have made a difference to the quality of teaching and learning." Hattie (2005) goes on to argue that, though USA spending on education had increased in the previous 40 years[7], the achievement curve remained constant over the same period of time. RSA is challenged with redressing the neglect of large portions of the education system during the apartheid era, but it is clear that this large investment in formal education alone is not improving the quality of learning in schools. This re-asserts the need to combine educational investment with appropriate monitoring at primary school level to improve the quality of teaching and learning.

[7] This increase in USA education spending remains, even if controlled for inflation. The figures can also not merely be downplayed to larger enrolment figures, as this increase remains, even if spending is viewed in terms of investment per learner and when costs for building new schools are factored out (Hood, 1990).

In RSA, educational data are collected through international comparative educational studies such as PIRLS (Howie, et al., 2008) and TIMSS 1995 (Howie, 1997), 1999 (Howie, 2001) and 2003 (Martin, et al., 2004). Systems level data are generated through systemic evaluations, which mirror the poor international performance. Poor performance is noted in both the Grade 3 and Grade 6 National Systemic Evaluation Reports (Department of Education, 2002b, 2006a, 2006b). School level monitoring is also mandated as part of the Internal Quality Management system of schools (IQMS) (Education Labour Relations Council, 2003). The mere availability of these data alone cannot improve learner performance, as the data also needs to be appropriately returned to schools and employed by them for planning, decision-making and action.

Feedback of data on learner performance to teachers and principals has long been regarded as generally enhancing performances of schools and learners (Coe, 2002). Kluger and DeNisis (1996) conducted a meta-analysis of 131 studies where feedback interventions were employed to bring about improvement in schools and they found that the average effect was moderately positive. However, more than 38% of the studies showed a negative effect of feedback of data and the mode of the effect size was zero. This research showed that, although feedback systems could have a positive effect at a school level, they do not have a consistently positive effect in all schools. Although this is the case, many countries around the world have turned to research and monitoring to improve the quality of education, teaching and learning (Jansen, 2001). Coe (2002) and Hattie (2005), amongst other authors, suggest that a great divide exists between monitoring and use of feedback of monitoring data in education to bring about improvement through praxis. Hattie (2005, p. 11) also notes that while there is an abundance of data collected and generated on schools the crucial question remains: “How can we return it to schools?”. In other words, how can the data be suitably fed back into schools to facilitate appropriate and constructive use? It seems that any data feedback system alone may not be the solution to improvement of educational provision, even if accompanied by increased

investment in this sector. It may be that a better quality of feedback system would make a difference here.

Feedback of data to schools is an important component required for improvement in education provision. Unfortunately, data disseminated to schools are often not used, or used inappropriately by the schools (Schildkamp & Teddlie, 2008; Wohlstetter, et al., 2008). This inappropriate use or non-use highlights the need for in-depth knowledge of the characteristics, conditions and dynamics that may enhance appropriate use of feedback based on data from school monitoring systems. It is this phenomenon of use of feedback that was investigated in this thesis and the knowledge thus gained may lead to improved educational delivery by the schools involved in the project. This knowledge may also provide a steppingstone to enrich approaches to performance[8] augmentation in education.

However, the question of how to enhance performance in schools is complicated by contextual factors. Some authors such as Fullan and Dalin (in Visscher, 2002, p. 52), state that because the situational factors vary greatly from school to school, no general system to feed data back to schools can lead to much benefit. As Fullan (in Visscher, 2002, p. 52) explains: "[t]here is no silver bullet", there is no solution which will fit all contexts.

The South African system is not only different from other educational contexts, but there is also great diversity within it. The learners come from multiple contexts with highly variable levels of skills and knowledge when they enter formal schooling. Once in formal schooling, they also develop at different rates. The severe shortage in qualified teachers further complicates the educational improvement efforts (Department of Education, 2006d). The system is also still grappling with the aftermath of apartheid and so is forced to deal with a myriad of social problems, with significant government spending in this sector still mostly aimed at rectifying the lack of infrastructure. While the

[8] Performance in this case is not only seen as the outcome of learner performance (as it is traditionally defined), but also of how the school fulfils its role as educational provider through evidence-based practice.

vast educational investment since 1994 has led to improvement of the educational infrastructure and enrolment, unfortunately, learner performance has not shown a concomitant improvement. This re-emphasises the need to combine infrastructural improvements with evidence-based practices[9]. The complexity of the context (see Chapter 3) makes developing an appropriate feedback system to facilitate use of data in schools a challenging task. Notwithstanding this complexity, the country cannot afford to be paralysed by all these obstacles. It is essential to develop a trustworthy, viable feedback approach to facilitate the use of learner performance monitoring, while remaining cognisant of contextual factors. The feedback system must be accompanied by design principles that allow the approach to be adapted to various contexts (The issue of analytical generalisability is explored in depth in Chapter 4). The stimulation of the appropriate use of learner performance data is an important building block in improving educational delivery and addressing inequalities.

To summarise, learner performance in South Africa is poor, whether measured internationally or systemically. The phenomenon that high levels of educational investment do not result in concomitant improved learner performance is observed not only internationally (Cassassus, 2001; Hayward & Hedge, 2005), but also in South Africa (World Bank, 2008; Department of Education, 2006a). Learner performance monitoring data for evidence-based practise are required to address the problem (Brinko, 1993; Hattie, 2005; Coe, 2002). The data can, however only have a positive impact if fed back to schools. Not all approaches to providing feedback, however, lead to improvement of educational delivery (Kluger & DeNisi, 1996). Schools often do not know how to use data appropriately, may not understand it or be unwilling to incorporate it in their decision-making process (Hattie, 2005; Schildkamp & Teddlie, 2008; Wohlstetter, et al., 2008). Contextual factors

[9] Controversy surrounds the issue of evidence-based practice in education with some associating it with a particular type of research associated with the Cochran Foundation. In this thesis, the term relates to adapting planning and educational practices in schools based on feedback of data from learner performance monitoring. The definition means pedagogical practices are interrogated based on the evidence of monitoring as feedback, and thus become more amendable to change based on this evidence. This is also sometimes referred to as data-driven decision-making, which relies on systematic analysis of data and application of the analysis (Schildkamp & Kuiper, 2009).

also play a large role in determining if a feedback system will succeed (Fullan & Dalin in Visscher, 2002, p. 52). This thesis attempts to address these issues by developing a contextually appropriate feedback system for learner performance data. The focus in development throughout was on facilitating understanding and use of the feedback by schools. Part of addressing this issue was exploring why data are not used or are used inappropriately. The thesis not only resulted in the optimising of a feedback system for the South African context, but also in design guidelines to support other designers in developing effective learner performance feedback systems.

Aims of the Study

The aim of this study was to identify and understand the characteristics of an effective feedback system and the utilisation thereof in order to design and optimise a feedback system that facilitates the use of learner performance data in South Africa within the school environment. The focus was on the management and classroom levels, specifically on principals, HoDs and teachers. This study of the feedback investigated use of both the processes of monitoring and feedback and the data generated by these processes.

The aim of the study was therefore two-fold:

1. To enhance, optimise and contextualise a learner performance feedback system
2. To identify design principles and characteristics of an effective learner performance feedback system

The first aim was achieved by using a design research approach to optimise an existing feedback system (SAMP). This was accomplished through gradual and successive cycles of design, implementation and evaluation of the feedback system prototypes. The focus of the design process shifted from contextualising the system, through establishing conditions for use of the feedback, to establishing finally how schools use the feedback.

The second aim was achieved by consulting existing literature surrounding the documented characteristics of an effective feedback system. This was

followed by an investigation into how to achieve the optimal conditions to allow for use of the feedback system in the South African context. The data were generated through evaluations of the feedback system prototypes. The processes schools employed to transform the feedback into planning and action in schools were also investigated, to adapt the feedback system to be appropriate for the current users. The process culminated in the development of design guidelines for an effective learner performance feedback system based on the literature review and data generated through the design research process.

The primary focus of the study is to enhance the feedback system and identify the associated design guidelines to facilitate use. This investigation includes both the identification of the optimum conditions for use of feedback that a monitoring facilitator should endeavour to establish, as well as the study of the processes in schools for transforming the feedback into action. The development of a feedback system and relevant design principles may cascade into change of facilitator, principal and teacher behaviour surrounding the use of feedback. The effects of this study may contribute to the improvement of teaching and learning in the wider South African context, as well as providing guidelines to feedback designers in different contexts.

1.4 Research Questions

The general research questions are presented in this section, to be operationalised in Chapter 3 and explored further based on the literature review and conceptual framework. The overall research question is:

- What are the characteristics of an effective feedback system and the use thereof in designing an optimum feedback system to facilitate appropriate use of learner performance monitoring in primary schools in South Africa?

Characteristics refer to the elements that should be present in such a feedback system, as well as how these should interact with each other in the context for which it was designed. The various characteristics form a gestalt where the interaction of the characteristics forming the whole is more than the

sum of the parts. An optimum feedback system, relevant to the context and needs of the users, must:

- achieve consistency between the various elements of the feedback system
- be practical for the users
- be effective in informing planning and practise for improving learner performance

A feedback system that facilitates use produces not only understandable and accessible data for schools, but also encourages and stimulates the use of the data for evidence-based practice. The learner performance monitoring in this study refers to data generated through the SAMP assessments on Grade 1 learners' performances in English Additional Language, Handwriting, Early Phonics, Early Reading and Early Mathematics. The definition of use incorporates not just instrumental or direct use of the data by schools, but also conceptual use or development of skills and approaches for working with data, planning and problem-solving that may result from participation in the feedback system.

In order to address the overall research question, it is necessary to examine a number of specific sub-questions. In this study there are six questions needed to answer the main question.

1. How can an existing learner performance monitoring system be appropriately adapted, contextualised and translated to the South African context?

The optimisation of the feedback system necessarily involves a parallel process of improving the implementation of the school-based monitoring system, SAMP. Although this improvement of SAMP is not the central focus of this study, it is necessary to establish that SAMP has been appropriately adapted and contextualised to the country to ensure it generates valid and reliable contextually appropriate data. The credibility of the assessment also contributes to the schools' perceptions of the data and likelihood that the

feedback will be used in the schools. Research sub-question 1 is addressed in Chapter 2.

2. What are the characteristics documented in literature of an optimal feedback system for use in school-based monitoring?

Existing knowledge reflected in the literature had to be incorporated in the design of the optimal feedback system for SAMP. Sub-question 2 helped to identify the essential characteristics of a feedback system for school-based monitoring as indicated by the literature. Further information about what is necessary for optimal use of the feedback was gained through the close study of other school-based monitoring systems that have led to effective use of data in the school environment. This study included an examination of the current South African educational context (see Chapter 3) in which the system functions. This information provided the basis for the global or overall design for the SAMP feedback system. Research sub-question 2 was the focus of Chapter 3.

3. What pre-existing conditions need to be established in the feedback system to facilitate the use of the learner performance feedback system?

Once the documented characteristics of an optimal feedback system were identified in the literature, it was necessary to establish how to optimise the various components of the SAMP feedback system. The primary focus was on investigating how the optimal conditions for use of the feedback system could be established. Research sub-question 3 was the focus of Chapters 5-8.

4. How do schools use feedback?

The purpose of the feedback system was not only to provide feedback to schools but also to facilitate use of the feedback for decision-making, planning and action in the schools. In order to understand fully how the feedback system could facilitate the use of the feedback in schools, it was important to study how feedback was being used in the schools. Research sub-question 4 was addressed in Chapters 7-8.

5. How effective is the feedback system in enhancing classroom practices, management and planning activities?

This question brings the study to the issue of praxis, or directed change and to what extent the feedback was transformed into action and, consequently, improvement in the school environment. In order for a feedback system to be effective, it must have a measurable impact on the quality of education in terms of classroom, management and planning activities in schools. In order to measure causal links in improvement in learner performance, an experimental design is required that has the potential to be an extensive study. Such a study is beyond the scope of this thesis, but evidence of changes in classroom, management and planning was documented through questionnaires, observations, interviews and monitoring data. This evaluation therefore includes both perceived efficacy as informed by input of principals, HoDs, teachers and academic consultants as well as some evidence of actual efficacy. Research sub-question 5 was addressed in Chapters7 and 8.

6. Which design guidelines for the development of an effective feedback intervention for school-based monitoring can be identified?

Finally, it is essential that there is a wider impact than this specific study and context. The design principles are required to identify the characteristics of an effective feedback system to address the main research question and contribute to the body of knowledge on school performance feedback systems. In order to provide a basis for transferability of this research to different contexts, design principles are identified from the design research process for use in other contexts. These principles may support other researchers in their own development of monitoring feedback systems that enhance evidence-based practice. Research sub-question 6 was addressed in Chapters 3 and 5-8.

1.5 Research Methodology

This research aims not only to generate knowledge by describing the characteristics of an effective feedback system and developing design guidelines, but also to design and develop a well functioning feedback system. The main research question lends itself to a design research

approach[10] that aims to align research and utility (De Villiers, 2005; Van den Akker, 1999). The design research process is iterative and follows a cyclical pathway of development (Nieveen, 1997; Richey, Klein, & Nelson, 1996; Thijs, 1999). As De Villiers (2005) explains, design research is a cyclical iterative analysis of design, development and implementation, combined with formative evaluation to understand the issues of the application domain. In this study, for example, each cycle of design research consists of the design and introduction of a version or prototype of the feedback system. This in turn is formatively evaluated, leading to a further cycle of development with a new prototype. Developing various prototypes is seen as generating "successive approximation of the ideals" (Van den Akker, 1999, p. 2). The prototypes are aimed at providing solutions in a real life context.

The prototypes in this study are gradual, successive approximations of the ideal feedback system for SAMP that encompass both the knowledge and the application domain. In this study four new prototypes were developed. Firstly, the feedback system design process was concerned with how to present and disseminate the data, as well as how to provide support to facilitate the understanding and interpretation of data by schools (Prototype I and Prototype II). Secondly, it was concerned with how to stimulate the appropriate use of the data for educational planning and application in schools (Prototype III). The final prototype (Prototype IV) along with the accompanying design principles based on the development process were the products of the study. Therefore, both the application and the research components were addressed.

The design research approach for this study incorporates various combinations of qualitative and quantitative methodologies during each evaluation cycle of the prototypes. As such, this study is located in the pragmatist paradigm, foregrounding issues of utility above those of method and propagating the use of the most appropriate tools to investigate a

[10] Design research was previously known as development research, as the research aims to design or develop a real world solution to a problem. Van den Akker (1999) in his earlier works referred to development research. Currently authors refer to design research (Plomp, 2009), while some prefer the term 'design-based research' (Joseph, 2004). In this thesis, the term 'design research' is used for consistency.

phenomenon (Onwuegbuzie & Johnson, 2004, 2006; Onwuegbuzie & Leech, 2005; Tashakkori & Teddlie, 1998). Pragmatism was a highly appropriate paradigm for this study as the study specifically focuses on the use of feedback (The choice of paradigm is discussed in full in Chapter 4). This design research study applied mixed methods, combining quantitative and qualitative approaches across all phases of the research, from conceptualisation to inference. Onwuegbuzie and Johnson (2006) describe this type of design as fully integrated mixed research and note that such an approach is attractive due to the multiple points of integration and complementarity it involves (The overall research methodology is discussed in Chapter 4).

1.6 Presentation Style

I have multiple roles in this research, including both monitoring and feedback facilitator and evaluator during the design research process. I explore these roles and implications for this research in depth in Section 4.3.1 and explain the role of self-reflexivity in managing these roles and the tensions.

Throughout the thesis some research diary and memo entries will be used for reflexivity and to make my disposition and process clear to the reader. Other than in these sections, I will refer to myself in the third person as 'the researcher' - a personal choice to facilitate the writing and reading process and provide consistency. This approach is appropriate since the research is a design research process and whilst I acknowledge my signature will have an influence on the research, it is not an emphasised component of design research.

1.7 Structure of this Thesis

The remaining chapters in this thesis are introduced below along with a short description of the content of each chapter.

- **Chapter 2: The South African Monitoring system for Primary schools**

This Chapter addresses research sub-question 1. The SAMP instruments provide the data and processes for which this study is developing and

improving the feedback prototypes. It is essential to have a firm grasp of the monitoring system itself. The aim of the chapter is to provide information about the development and South African contextualisation of the instruments. It concludes with a description of the quality of the monitoring data that are generated by the SAMP instruments.

- **Chapter 3: Contextualisation, literature review and conceptual framework**

This chapter addresses research sub-question 2. The feedback prototypes for this research were designed for the South African schooling system and policy context. The chapter starts by describing the country's education system and concludes with design criteria for the design of the prototypes in that context. This chapter also provides a rich description of the South African context to afford the readers an opportunity to transfer or adapt this research to their own contexts. Four prominent international school information systems are described. This provides a steppingstone to the development of the feedback prototypes. The chapter examines literature from the fields of monitoring utilisation, feedback, school improvement and effectiveness to develop the conceptual framework that guided the study and interpretation of the findings.

- **Chapter 4: Overall Research Design**

This chapter starts with a discussion of the ontological, epistemological and methodological basis of this study. The design research approach applied in this study is explored. The chapter concludes with a discussion of the methodological norms of the study and ethical considerations. Only an overview of the research design and the choice of the research approach are discussed in this chapter. Detailed descriptions of the design and methods for each cycle are provided in Chapters 5-8 as each successive cycle is discussed.

- **Chapter 5: Preliminary Phase: Problem identification, needs and context analysis**

This chapter focuses on research sub-question 3. The aim is to examine the problem in context and establish the criteria for an optimal feedback system. The chapter explores an exemplary case study of a proven and effective feedback system, namely the asTTle (assessment tools for teaching and learning) in New Zealand (NZ). The NZ context was compared to the South African context and design criteria for the SAMP feedback system were generated. A detailed discussion of the research procedures for this cycle is combined with the presentation of the data and design principles.

- **Chapter 6: Prototyping Phase: Establishing conditions for use (Cycle 1-2)**

This chapter documents and discusses the design, procedures, data and results for the first two design cycles. Sub-questions 3-4 are the focus of this chapter. The cycles aimed to establish conditions for the use of the feedback and launched the investigation into how schools used the data. Data collection included expert evaluations, the Delphi technique and questionnaires. Two design prototypes were developed and evaluated. Specific sampling, analysis and research procedures for the two cycles are explored. The chapter explores the evolution of the design principles from one feedback prototype to the next.

- **Chapter 7: Prototyping Phase: Transforming conditions for use (Cycle 3)**

This chapter documents and discusses the design, procedures, data and results for the third design cycle. The focus of this cycle is on sub-questions 4 and 5, although elements of sub-question 3 are also expanded. The cycle aims to establish how conditions of use are transformed into classroom, planning and management practices in schools. Data collection included questionnaires, observations, structured reflective journals and interviews. Specific sampling, analysis and research procedures for the cycle are explored.

- **Chapter 8: Assessment Phase (Cycle 4)**

This chapter documents the semi-summative evaluation of the fourth and final prototype for this thesis. The design, procedures, data and results for the cycle are discussed. The focus of this cycle is sub-questions 6, developing the design guidelines for the study. Sub-questions 2-4 are also explored for the final prototype. Data for the semi-summative evaluation were collected through expert evaluation reports and questionnaires for teachers and school management.

- **Chapter 9: Conclusions and recommendations**

The final chapter presents a summation of the findings of the research and the conclusions to be drawn from this study. In particular, it explores the implications of the design principles for the SAMP feedback system as well as the applicability of these findings to other contexts. A set of conclusions and recommendations is presented, along with a discussion of the possible effects of these findings on policy, practice and research. The chapter therefore addresses the overall research question. The limitations of this study are explored along with recommendations for further research.

CHAPTER TWO

The South African Monitoring system for Primary Schools

> An implicit assumption of almost all school effectiveness research is not only that pupil progress is good but that the schools that encourage the most rapid progress are the best. This largely untested assumption seems most likely to hit problems for very young children starting school. (Tymms, et al., 2000, p. 105)

The first research sub-question is addressed in this chapter: How can an existing learner performance monitoring system be appropriately adapted, contextualised and translated to the South African context? The chapter provides information on how the original PIPS monitoring instruments were developed in the UK, then translated, and contextualised for South Africa to become SAMP. The adaptation, translation and contextualisation of the Vocabulary subtest of the PIPS instruments are used to illustrate this process. The data quality of SAMP in the South African context is explored. The validity and reliability of data produced for the sample of schools through the SAMP instruments are established to show that the data could be used with confidence to examine use of the feedback system in the South African context.

2.1 The PIPS Instrument

The CEM was established in 1983, based at the University of Durham, it is the largest university-based educational research unit in the UK (Curriculum Evaluation and Manangement Centre, 2007). When the National Curriculum was introduced in the UK in 1988, it led to a formidable public accountability system for English schools. The accountability system ranges from the publishing of league tables to regular inspections by the Office for Standards in Education (OFSTED). The OFSTED findings are also published and made available to the public on an OFSTED website (Tymms & Albone, 2002). The CEM developed out of a drive to improve the education system from within i.e.

changing accountability practises to be more improvement-orientated through the existing structure.

The CEM is rooted in the concepts of 'distributed research', ownership and participation by practitioners to solve educational problems (Fitz-Gibbon, 1996; Tymms & Coe, 2003). The CEM's professional monitoring systems are for use by practitioners and not meant as official accountability systems that may lead to blaming of practitioners (Tymms & Coe, 2003).

The CEM has developed a number of monitoring systems that are used by about 7,000 schools in the UK and assess the progress made by over a million learners every year (Tymms & Coe, 2003). The CEM suite of monitoring systems (see Table 2.1) caters for learners from 3-19 years (Curriculum Evaluation and Manangement Centre, 2007; Tymms & Coe, 2003). This suite of value-added assessments was developed in conjunction with teachers and Local Education Authorities (LEAs) staff in the UK. The assessments combine objective assessments and teacher ratings to provide valuable information about each learner (Tymms & Albone, 2002).

For the South African adaptation, two points in schooling were identified as crucial areas where monitoring was necessary, namely entries into primary and secondary school (Howie, 2002). South African schools often receive learners with highly diverse backgrounds and levels of skills from wide feeder areas when entering primary and secondary education (Scherman, et al., 2006). At the time, there were also no South African monitoring systems in place that focused specifically on these transitional points. This meant adapting the CEM's assessments for the beginning of primary school (Performance Indicators in Primary Schools or PIPS) and the beginning of secondary school (Middle Years Information System or MidYIS).

Table 2.1: CEM information systems

ASSESSMENT	AGES	FOCUS
A Level Information System (ALIS)	16-19 years	Vocational and academic Value-added measures up to Advanced level examinations.
Year 11 Information System (YELLIS)	14-16 years	The baseline for GCSE grades, comprising mathematics, vocabulary and perceptual reasoning. Value-added scores can be calculated and additional information captured.
Middle Years Information System (MidYIS)	11-16 years	This is a curriculum free assessment examining mathematics, vocabulary perceptual reasoning, proofreading and perceptual speed and accuracy.
Performance Indicators in Primary Schools (PIPS)	5-12 years Infant & Junior	Provides assessment for all primary schools, used in England and Scotland. Examines developed abilities in vocabulary and non-verbal ability as well as outcome measures in terms of mathematics and reading in order to provide concurrent value-added measures.
Performance Indicators in Primary Schools (PIPS)	4-5 years On-entry baseline and follow-up	Assesses variables shown to be good predictors of later success in schooling. Value-added is measured against Foundation Phase profile at end of reception in England.
Interactive Computer Assessment System (InCAS)	5-10 years	A personalised diagnostic assessment as part of the monitoring systems
Assessment Profile on Entry for Children and Toddlers (ASPECTS)	3-4 years	Prevent age-related feedback compiled from information from home and preschool as well as assessments of physical personal and social development as well as language and mathematics.

(Curriculum Evaluation and Manangement Centre, 2007; Tymms & Coe, 2003)

This PIPS instrument adapted for South Africa provided the basis for the development and evaluation of the feedback system prototypes in this thesis. The CEA decided to adopt PIPS, adapting it for the country's context and implementing it by means of funding from the NRF (Howie, 2002). The PIPS assessment fulfilled the CEA's criteria for an assessment measure in that it provided an indication of a child's readiness for academic learning as scores on the test administered at the start of schooling and correlated well with subsequent academic achievement (Tymms & Coe, 2003). PIPS was also administered twice a year to provide a measure of progress and the

assessment provided information on a child's profile of performance in a number of domains that could be used to identify particular learning difficulties or strengths.

PIPS was developed in 1994 for the purpose of providing a baseline by which to assess progress in reading, phonics and mathematics for learners entering primary school (Tymms, et al., 2004). The monitoring system aimed to improve education through feedback to the schools (Tymms, 1999). The baseline assessment was combined with a follow-up assessment to determine the value added[11] by schools between the two assessments.

The full PIPS instrument consists of 14 subtests, 13 of which are combined into three different scales: Early Phonics, Early Reading and Early Mathematics. The Handwriting subtest is separate and constitutes a scale score on its own. The scales are generated as follows:

1. The ***Early Phonics*** scale is derived from two subtests, the Repeating Words and Rhyming Words subtests. These subtests focus on phonic awareness as an important basis for the development of reading ability.
2. The ***Early Reading*** scale is derived from seven subtests: Vocabulary, Ideas about Reading, Letters, Words, Story and Sentences. This scale focuses on the prerequisite skills for reading development.
3. The ***Early Mathematics*** scale comprises seven subtests, namely Sizes, Counting, Sums A, Numbers and Mathematics. These subtests aim to establish the learners' abilities in early mathematics.

The baseline assessment consists of only 11 subtests, which are repeated in the follow-up assessment, during which a further three more advanced subtests are also administered. The results from the follow-up and baseline assessment are compared to determine value-added scores between the two assessments.

[11] Value-added measures quantify the value added by a school to learning for a learner, taking in consideration various factors. There are different ways to calculate value-added scores, these are discussed in depth in Section 2.2.1

PIPS was originally developed as a paper-based test, but in the UK is usually employed in its computer-delivered format (Merrell & Tymms, 2005). The PIPS assessment was administered individually and usually took 20 minutes per learner. Tymms and Wylde (2003) report that schools generally find the administration time manageable.

Following the widespread success of PIPS in the UK, it was adapted with minimal changes for countries such as Australia, New Zealand and Scotland. The Dutch and British Sign Language versions were developed to maintain the original intentions and characteristics of the items (Merrell & Tymms, 2005). Germany, Lesotho, Thailand, France and Hong Kong have also adapted PIPS (Tymms, et al., 2004). The CEA is responsible for the South African adaptation of PIPS known as SAMP.

With PIPS being implemented in several countries, the opportunity for international comparison of school entry levels of learners was created (Merrell & Tymms, 2005). The use of the data for international comparative purposes however reiterated the importance of appropriate adaptation that maintained the difficulty levels and intentions of items across applications in various countries. As the South African context differs widely from that in the UK, the unique learning context of South Africa was expected to influence how children perform on the CEM instruments. Therefore, it was necessary to adapt aspects of the monitoring system.

2.2 The South African Birth of SAMP

A number of effective learner performance monitoring systems exist (see Chapter 3 for a discussion of four such systems). The CEM instruments were chosen for adaptation as the CEM expressed interest in working with the CEA and in fostering monitoring and evaluation skills in South Africa. The CEM was also willing to provide their instruments to be adapted and implemented in the country without charge and to provide support where necessary during development. The CEM suite of instruments has also enjoyed widespread sustained success and has been adaptable to various international contexts, whilst maintaining its integrity.

The CEA and the CEM started to collaborate in 2003. The CEA decided to adapt PIPS and MidYIS for the South African context and implement them at the beginning of primary and secondary school. Funding was provided by the NRF.

The project aims to provide quality data to schools. Performance data are aggregated on school, class and learner levels allowing for monitoring of learner performance and improvement of educational provision by schools. The system enables schools to monitor their own performance outside of a formalised accountability framework. These data may also be used to inform policy development. The initial research questions to be addressed by the project were as follows:

1. How feasible is a monitoring system using value-added measures in South Africa?
2. How valid and reliable are existing value-added measures for South African education?
3. What additional assessments will need to be developed to enhance monitoring systems?
4. What is an effective means of providing feedback to schools using value-added measures?
5. What assistance will the educators need in using the results from the assessments?
6. How will the effectiveness of the monitoring system be evaluated?

This thesis extends the original project's research question four. The focus here is not only on the most effective way to provide feedback, but also on how to provide support to understand the feedback and to transform it into evidence-based action in the school environment.

The primary school project (PIPS) was brought to South Africa for a number of specific purposes. The first was to provide an indication of a child's readiness for academic learning. Scores on the test administered at the start of schooling correlate well with subsequent academic achievement (Tymms &

Coe, 2003). Secondly, the same test is administered again at the end of the first year at school to provide a measure of progress made. Thirdly, baseline scores can also be used to estimate progress in further years. Fourthly, the test provides information on a child's profile of performance in a number of domains that can be used diagnostically to identify particular learning difficulties or strengths. Fourthly, the data from the instrument can be used to adapt the classroom teaching practises to suit the needs of the group and individual learners.

Using a value-added approach to measure learner performance is highly appropriate for the South African context as it considers prior performance. The approach is more equitable and provides for a more constructive feedback than data that ignores learners' baseline skills, especially in a country where great learner diversity has historically been employed as the basis for discrimination. The importance of the value-added approach in monitoring learner performance is discussed in the following section.

2.2.1 The Importance of Value-Added Measures

As the term implies, value-added measures assess the value that a school adds to learner achievement while considering the influence of intake factors such as background, prior achievement, aptitude and abilities (Scheerens, et al., 2003). Value-added measures are generally seen as a more equitable and balanced approach to monitoring of learner and school performance when compared to measures which ignore these factors (Rowe, Turner, & Lane, 2002; Saunders, 2001). Value-added measures engage a variety of statistical methods to adjust gross output indicators to incorporate intake factors (Scheerens, et al., 2003). Bosker and Witzier (in Rowe, et al., 2002, p. 172) differentiate between three different kinds of value-added measures:

1. **Unpredicted achievement:** adjusted for family background factors and student ability.
2. **Learning gain:** adjusted for initial achievement level.
3. **Net progress:** adjusted for family background factors, ability and initial achievement.

Currently the SAMP project uses the learning gain measures for schools in the sample. The monitoring system allows for easy comparison of baseline and follow-up performance results that is necessary for a learning gains approach.

The importance of value-added measures can be illustrated by examining the factors that contribute to learner performance. As Saunders (2001) explains, raw scores reveal more about student background than about school performance. In Figure 2.1, Hattie (2005, p. 13) illustrates the role players and their relative contributions to variance in learner performance.

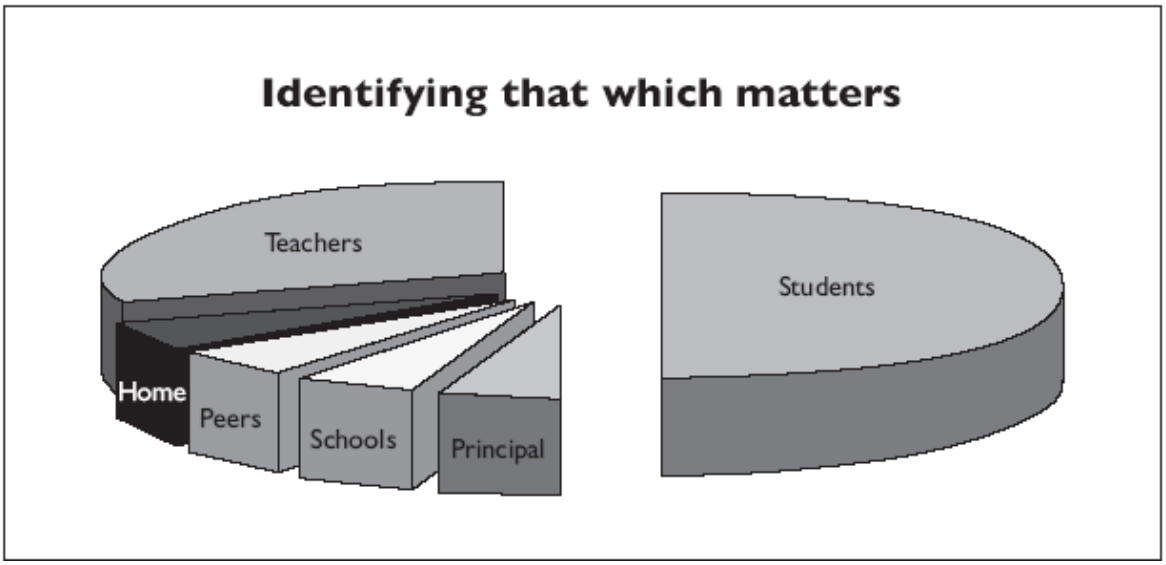

Figure 2.1: Contributions to learners' achievement variance
(Hattie, 2005, p. 13)

As the pie chart indicates, factors inherent in the learner (e.g. intelligence, maturity, socialisation, motivation and aptitude) are responsible for 50% of the variance in learner performance. Systems such as the league table published in the UK neglect intake factors inherent in the learner, which could misrepresent schools. Omitting learner intake characteristics creates the impression that the intake learner profiles of all the schools are similar or irrelevant and that any ranking is thus purely based on the quality of the school. This type of accountability system may have multiple adverse consequences:

- Political and media 'bashing' of schools and teachers
- Test-dominant curricula
- Schools selecting stronger pupils for intake to boost 'league table' rankings
- Parents taking drastic measures to enrol their children in high ranking schools

(Rowe, Turner, & Lane, 2002, p. 166)

Schools in the new democratic South Africa have also experienced media persecution. This could be seen with the listing in the press by the DoE of the worst schools in each of the nine provinces (Jansen, 2001), a practise that was only ceased under the guidance of Minister Naledi Pandor. Some countries have taken a very firm stance to prevent public comparison of schools. These countries include Ireland that introduced legislation in 1998 to prevent school-by-school comparison (Looney, 2006).

In South Africa, the need for value-added measures - as opposed to performance outcome measures alone - is accentuated by the great diversity amongst learners. Factors and conditions that influence learner performance range from issues of language and socio-economic circumstances to an educational system that is still trying to align schools under one education department (schools were categorised under various education departments during the apartheid era). Any monitoring system that remains blind to these factors will not generate quality data for feedback to schools in the South African context. The next section examines the adaptation of the PIPS assessment to show how it was contextualised and translated to the South African context to become SAMP.

2.2.2 Contextualisation and Adaptation of SAMP

The adaptation and contextualisation discussed here are illustrated through examples. In this case, the Vocabulary subtest items are used to illustrate how SAMP was adapted and contextualised. The complete, extensive process is detailed in Archer, Scherman, Coe and Howie (2010).

The first step in contextualising PIPS was to determine the aspects of PIPS that presented difficulty in the South African context and needed to be re-examined (see Section 2.2.2.1). The second step was generating suggestions for adapting the identified problematic items (see Section 2.2.2.2). Thirdly, the functioning of the newly adapted instrument had to be assessed (see Section 2.2.2.3). The vocabulary subtest is used throughout the discussion as the illustrative subtest. The vocabulary subtest was selected as it was subjected to multiple graphical and language changes during the adaptation.

2.2.2.1 The PIPS Instrument in South Africa Prior to 2006

In 2003, the UK English language version of the PIPS instrument was translated and adapted for use in three South African languages[12]: Sepedi, Afrikaans and English. These three languages represent the three dominant LOLT in the Tshwane region, the region in which the CEA is located (H. Julies, DoE, personal communication, July 16, 2010). The original UK-based PIPS instrument was translated into Sepedi and Afrikaans by registered translators, then corroborated through a process of back-translations. The translated instructions and items were recorded for the computer-based assessment by voice artists and the voice clips were sent to the CEM to be incorporated into the code of the computer-based PIPS. By 2005, the PIPS computer-based assessment was available in all three languages with learners guided through the assessment on a one-on-one basis by a fieldworker who captured answers directly on the computer as the learner provided them.

When this researcher was appointed to coordinate the project in 2006, 416 Grade 1 learners were participating in the project. This included learners from two Afrikaans, two Sepedi and three English Language of Learning and Teaching (LOLT) schools in the Tshwane area. A team of seven trained fieldworkers would travel to each school with rented laptop computers and the software to assess the Grade 1 learners. At the time, sustainability, growth

[12] All learners are assessed in the LOLT of the school and not their home languages as the assessment aims to establish readiness and ability to perform in their current academic setting. This means that many of the learners are not first language learners. The most diverse learner population in terms of home languages is found in the English LOLT schools.

potential and the cost of the project were of concern. A review of the progress of the project and planning for it further development was conducted.

The review of the project status (for full reports see Archer, 2006a, 2006b) examined the handover and process documentation from the previous project team (Eiselen, 2004; 2005b; Küstner & Eiselen, 2005) and a report by an educational psychologist (Barry, 2005) on the cultural fairness of the instrument. It soon emerged that administering the assessment on computers contributed disproportionately to the project expenses and both slowed and complicated adapting and contextualising the instruments. Theoretically, the computer-based assessment offered a number of benefits over paper-based assessment:

1. **Ease of capturing of data:** Data are directly captured on the computer during administration.
2. **Ease of administration:** In its computer-based format, the PIPS assessment automatically implements termination rules. The fieldworker is guided through all the administration procedures in a systematic fashion and the fieldworker is forced to capture the data before the next item is revealed.
3. **Interactive nature of the presentation:** As a computer-based assessment, the learners are confronted with and interactive environment. Murphy and Davidshofer (1994, p. 191) state, "[t]here is evidence that people view computerized tests more positively than equivalent paper-and pencil tests. This often contributes to the learner's ability to maintain concentration and interest in the assessment".
4. **Facilitation of international comparability:** PIPS is administered as a computer-based assessment in most of the participating countries. Computer administration thus facilitates comparison of the South African data with that of other countries. (Archer, 2006a)

These perceived benefits of computer–based administration of PIPS were not realised in the field in South Africa. The CEA was obliged to reconcile the perceived benefits of computer-based assessment with the realities of the country:

1. **Ease of capturing data**: Several difficulties with capturing the data on the laptop computers in the field were discovered and reported by Küstner and Eiselen (2005). These difficulties included incorrect entries that were difficult to correct due to an absence of hardcopies of the assessments.

2. **Ease of administration:** Many of the fieldworkers did not have the required level of proficiency to utilise a computer. Extensive training in use of the computer-based assessment was required.

3. **Faulty termination rule application:** It was found that the computer programme would sometimes terminate pre-maturely irrespective of the learner's responses. (This can be seen in the Letter subtest data where the test terminated after eight items irrespective of learner responses). It was also difficult to check for such errors, as the CEA did not have direct access to the programming code for the assessments.

4. **Incorrect insertion of sound clips**: As the PIPS assessment remains the property of CEM, the translated sound clips for the assessment in Afrikaans and Sepedi had to be sent to CEM to be inserted into the programme. This resulted in some of the clips being placed incorrectly.

5. **Cost of administration:** Most of the schools involved in this project did not have computer laboratories. Therefore, laptop computers had to be rented for the fieldwork. The cost of renting laptop computers for the fieldwork represented a major part of the expenditure in the project.

6. **Security:** Travelling with valuable equipment such as laptops presents a serious security risk in South Africa. This negatively affected the safety of the fieldworkers.

7. **Administrative burden:** The process of booking, renting, collecting and returning the laptop computers, as well as having to upload the necessary software repeatedly for fieldwork was a large administrative burden to the CEA team.

8. **Administration time:** The administration time of twenty minutes projected per child (Tymms & Wylde, 2003, ¶ 16) was greatly increased in the PIPSSA project, as laptops had to be set up and fieldworkers were often not as computer literate as the teachers in the UK are.

9. **Sustainability:** In order to achieve true sustainability for this project it would be necessary to empower teachers to administer this test and relay the data to the CEA. In order to achieve this, it is essential to ensure that the necessary infrastructure is in-place. Currently there are vast discrepancies in the availability of computer facilities for schools in South Africa. The South African Department of Education Draft White Paper on e-education of August 2003 indicates that in 2002 only 26.5% of schools had access to computers for teaching and learning (Department of Education, 2003b). Thus, the computer-based assessment would prevent certain schools from administering the assessment themselves.

10. **Adaptation for cultural fairness:** One of the CEA's objectives with the PIPPSA project was to adapt PIPS to be culturally fair for the South African context. In order to do this certain items had to be amended. Items need to be translated and various other changes were required. As programming can only be done at the CEM this entailed a cumbersome process of negotiating changes and sending changes to be programmed in the UK and then be sent back.
(Archer, 2006a)

Based on these findings, the decision was made to change from computer-based to paper-based administration. A new name, SAMP, was adopted to signify this shift. As part of the transition, the equivalence of two types of administration had to be established. The equivalence was investigated through a small sample of 96 learners. All learners were assessed twice, using the paper-based and then the computer-based formats. Half of the learners were assessed using the computer-based format first and the other half with the paper-based first to control for any learning effect.

No significant difference was found in the performance of learners overall or on any of the scales, other than the Early Reading Scale (See Table 2.2). Particular note was made of the subtests in which there was a significant difference between the two modes of administration for further investigation (Vocabulary, Ideas about Reading, Ideas about Maths and Sums A). All these subtest were heavily reliant on graphic elements and enjoyed particular attention during the adaptation process discussed below. Although the sample was small, it provided sufficient information to determine that the two modes of assessment were fairly equivalent for the selected sample. Greater equivalence would be achieved through the adaptation process. These results re-affirmed the feasibility of moving forward with the development and adaptation of the SAMP project in paper-based format.

Table 2.2: Difference between paper-based and computer-based modes of delivery of the PIPSSSA assessment

SUBTEST/SCALE	PAPER-BASED	COMPUTER-BASED	DIFFERENCE	SIGNIFICANT AT 0.1 LEVEL
Vocabulary	33.5	30.5	3.0	*
Ideas about Reading	29.0	41.0	-12.0	*
Repeating Words	66.9	66.3	0.5	
Rhyming Words	28.9	31.2	-2.3	
Letters	16.7	15.5	1.2	
Words	8.7	12.0	-3.3	
Ideas about Maths	73.0	79.8	-6.8	*
Counting	78.0	81.3	-3.3	
Sums A	54.2	48.1	6.1	*
Numbers	32.1	32.9	-0.8	
Early Phonics Scale	47.9	48.7	-0.9	
Early Reading Scale	22.0	24.7	-2.8	*
Early Mathematics Scale	59.3	60.5	-1.2	
Total Scale	43.0	44.7	-1.6	

2.2.2.2 Adaptation of PIPS into SAMP during 2006

Establishing the validity and reliability of the instrument was key to determining how well it functioned in South Africa. When adapting an instrument, the core validity issue is determining what adaptations and accommodations preserve the meaningfulness of the scores (Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000). Validity is achieved through the removal of irrelevant construct variance, created by the difference in culture, context, language, and social practices. Therefore, an important aspect of establishing the 'validity argument' (Kane, 2006) is to demonstrate that test scores measure the same thing across all groups for whom the instrument is intended.

The core validity issue in adapting an assessment for the South African context is therefore determining which adaptations and accommodations would preserve the meaningfulness of the scores (Fuchs, et al., 2000). Validity addresses the question as to what extent the interpretation of results is appropriate as well as meaningful (Gronlund, 1998). Validity is a unitary concept that is based on various forms of evidence, with construct-related validity being the central concept. Ultimately, validity is concerned with the consequences of using the assessment (Gronlund, 1998; Killen, 2003; Linn & Gronlund, 2000).

One strategy for identifying bias in an assessment is to look for differential item functioning (DIF) (Smith, 2004). If the relative difficulty of an item differs significantly across various groups, it indicates that scores that include that item are not measuring a uni-dimensional construct. This means that performance on the item is being influenced by some characteristic of that group other than the underlying construct being assessed (Smith, 2004).

Generally, reliability refers to the consistency of scores obtained by the same individuals when they are requested to complete the assessment on different occasions (Anastasi & Urbina, 1997). Reliability not only indicates how much confidence can be placed in a particular score, but also how constant the scores will be over different administrations (Owen & Taljaard, 1996). Reliability for the assessment was established by investigating internal consistency, which is seen as a pre-requisite for construct validity. High inter-item correlations are expected among items that measure the same construct (Kline, 1993).

Reliability of Subtests and Scales

The Cronbach's alpha values for the 2005 computer-based assessment (see Table 2.3) indicate that some aspects needed to be re-examined. The acceptable level for a reliability figure is determined in part by the envisaged use of the data from the assessment. If the data are to be used for decision-making for a group, the figure can be lower than when it is used for decision-making concerning individuals. In the same way, a lower reliability figure can be tolerated if the data are not used in isolation, but in conjunction with other data (Frisbee, 1988).

Table 2.3: Reliability coefficients for the computer-based 2005 PIPSSA subtests overall and according to language of assessment

SUB-TEST / SCALE	OVERALL (n = 417)	AFRIKAANS (n=62)	ENGLISH (n=211)	SEPEDI (n=144)
Vocabulary	.85	.92	.85	.63#
Ideas about Reading	.87	.88	.88	.81
Repeating Words	.84	.88	.77	.85
Rhyming Words	.86	.84	.83	.77
Letters[13]	-	-	-	-
Words	.93	.93	.89	.86
Ideas about Maths	.83	.95	.75	.87
Counting	.85	.90	.80	.89
Sums A	.88	.87	.88	.84
Numbers	.91	.92	.90	.90
Early Phonics Scale	.85	.86	.80	.79
Early Reading Scale	.92	.92	.91	.81
Early Mathematics Scale	.93	.95	.91	.91
Total Scale	.96	.95	.95	.91

- items to be investigated further as indicated by Cronbach's alpha values

For the PIPPSA assessment, reliability values of above 0.8 were aimed for. It should be noted that creating reliable assessments for very young children is notoriously difficult (Archer, et al., 2010). Over and above indicating the stability of measures over time, a high reliability figure would also strengthen the inferences made about the content-related validity of the assessment (Suen, 1990). The reliability figures for the 2005 data were encouraging, however the low reliability figure of 0.63 for Sepedi learners on the Vocabulary subtest was of concern and highlighted the need to investigate the subtest further[14]. It was necessary to determine if the low figure was related to the construction of the subtest overall, the graphical presentations used, or, for instance specific phrasing of items in the subtest. As validity is related to whether the assessment measures what is intended to, the aim of the subtest is pivotal. The vocabulary subtest aims to evaluate the receptive vocabulary of learners and consists of 23 items. Learners are asked to point out objects in three different pictures, graded according to difficulty of the visual stimuli and items:

[13] No results for the Letters subtest in 2005 are available as the test terminated prematurely after only eight items.
[14] All other subtests for all three language groups were investigated further through the means described below, the Vocabulary items are used to illustrate the process here.

1. Kitchen scene (easy items) e.g. “Can you point to some carrots?”
2. Outdoor scene (moderate items) e.g. “Can you point to a windmill?”
3. Toy shop (more advanced items – examines learners’ exposure to literature) e.g. “Can you point to a yacht?”

A termination rule is applied. The rule requires that the subtest be discontinued when a candidate supplies three consecutive incorrect answers.

The report from Barry (2005) provided some indication that the computer-based PIPS was likely to disadvantage learners who had not been exposed to cartoons, animations or three-dimensional overlays. This may have been the source of the discrepancy in reliability figures for the Sepedi learners, as they were mostly from relatively poor socio-economic areas. The individual item statistics needed to be examined to determine if the graphic presentation alone explained the low reliability for Sepedi learners, or if the individual items also needed to be revised. The item facility and discrimination values for the subtests were studied more closely through classical test theory.

Item Facility and Discrimination Values

The item facility (also referred to as ‘item difficulty’ or ‘difficulty values’) and item discrimination values were used as indicators of items that needed closer examination. Item discrimination indicates the ability of an item to differentiate between high and low achievers. Item discrimination values of 0.25 or higher were aimed for when examining the item-total correlation values. Facility values show the percentage of learners who correctly answered the items. These values are presented separately for the first two pictures used in the Vocabulary subtest (see Table 2.4). As the termination rule was applied in the vocabulary subtest, most candidates were not presented with the most difficult items in picture 3. For the purposes of calculating item facilities, these missing items are treated as incorrect. Only the first 17 items are discussed here, for illustrative purposes.

Table 2.4: Difficulty and discrimination values for items overall and across languages

ITEM	ENGLISH			AFRIKAANS			SEPEDI		
	n	Facility	Discr	n	Facility	Discr	n	Facility	Discr
1- carrots#	211	92.9	0.23#	62	98.4	0.21#	144	82.6	0.39
2 - the knife#	211	79.6	0.39	62	98.4	0.21#	144	93.8	0.44
3 - a fork#	211	90.0	0.31	62	98.4	0.21#	144	97.2	0.44
4 - a cupboard#	207	72.0	0.35	61	90.3	0.18#	141	63.2	0.33
5 - some cherries#	198	59.2	0.36	61	61.3	-0.19#	141	59.7	0.25
6 - a pan	197	63.0	0.46	61	93.5	0.28	141	91.7	0.31
7 - a bowl#	183	48.8	0.32	59	58.1	-0.14#	139	75.7	0.33
8 - the butterfly	159	12.8	0.57	57	17.7	0.89	132	2.1	0.33
9 - the kite	140	11.8	0.58	54	17.7	0.78	127	4.9	0.29
10 - the castle	112	7.1	0.67	44	19.4	0.85	110	0	0
11 - the wasp	39	4.3	0.62	14	19.4	0.85	10	-	-
12 - the pigeon	26	5.7	0.61	13	14.5	0.82	5	-	-
13 - the windmill	17	3.3	0.66	12	14.5	0.91	3	-	-
14 - the turtle	14	5.2	0.70	12	17.7	0.89	3	-	-
15 - the violin	13	2.8	0.64	10	12.9	0.81	3	-	-
16 - the padlock	11	2.8	0.62	11	12.9	0.85	1	-	-
17 - the toadstool	11	0.9	0.28	9	9.7	-	1	-	-

- items to be investigated further as indicated by discrimination values

For very easy items (high facility value) such as item 1, the discrimination value is lower than the identified 0.25. This is to be expected as both low and high achieving learners typically answered correctly. The item could therefore only differentiate between very low performing learners and other learners. These items were maintained, as it is necessary to make allowances for very low performance learners to achieve a sense of accomplishment and to allow learners to become comfortable with the assessment.

From the item discrimination values, it seems that two items had a negative item discrimination value for the Afrikaans learners:

- Item 5 - Can you show me some cherries?

The Afrikaans translation of cherries is *kersies,* which is a homonym for birthday cake candles and cherries, both of which appear in the picture. This meant that learners often indicated the birthday candles instead of the more difficult item of cherries. Fieldworkers indicated that the same problem

occurred with many Sepedi learners as words in Sepedi are often borrowed from Afrikaans.

- Item 7 – Can you show me a bowl?

The approved translation of bowl as *papbakkie* for Item 7 of the Afrikaans assessment seemed to be problematic and an alternative, simpler translation was suggested to address the negative discrimination value.

The fact that most of the items were significantly more difficult for the Sepedi learners than English or Afrikaans learners is of concern. This may be due to the way items have been translated or graphically represented, or it may be that whilst translations are accurate, the translations are less frequently used or are more advanced words that decrease the item facility. If the graphic representations are found alien or distracting by the Sepedi learner, they may well act as confounding variables in measuring receptive vocabulary. Alternatively, the particular sample of Sepedi learners tested may have a poor vocabulary. It was necessary to explore this phenomenon further, through techniques such as Rasch analysis.

Rasch Analysis

The Rasch modelling locates both the difficulty of items and the ability of persons on a single latent trait continuum. The probability that a person of ability, β, will correctly answer an item of difficulty, δ, is entirely determined by the difference β – δ. This means that the relative difficulty of two items on the continuum is independent of the abilities of the sample of persons who have attempted them (Baker, 2001). This is a particular strength in the South African context, where the aim is to examine Differential Item Functioning (DIF) across groups whose average scores are quite different. The aim here was to establish if the *relative difficulty* of the items were similar across the three language groups. The difference in *average performance* of the three groups becomes irrelevant as only the relative difficulties of the items to each other for each group is examined. The assumption is that similar relative difficulty values should be obtained across the languages to show equivalence of the instrument across the languages.

All 23 items from the three pictures of the Vocabulary subtest were included in the Rasch analysis of the Vocabulary subtests. Only items from the first two pictures, items 1-17, are discussed here to illustrate the process. Rasch analysis copes well with the missing data, this is beneficial in examining assessments where a termination rule is applied.

A person separation reliability of 0.67 was achieved in the assessment, indicating that the scale discriminates well between persons. The items also created a well-defined variable or single underlying construct (as indicated by the item separation reliability of 0.98). The OUTFIT mean square for both persons and items were slightly more than one, indicating underfit (1.04 and 1.57 respectively). Conversely, the INFIT mean square for both persons and items were below one, indicating overfit, or that the responses are too predictable (0.83 and 0.80 respectively).

Upon inspection of the items, it was found that the point-measure correlations were all positive and above 0.3 for every item included in the analysis. However, several items are identified by misfit statistics (namely Item 4 *'Can you point to a cupboard?'* Item 5 *'Can you point to some cherries?'* and Item 7 *'Can you point to a bowl?'*. These three items, although falling within the criteria of 0.5-1.5 for productive items with regard to the INFIT mean square, did not fall within the prescribed range for the OUTFIT mean square indicating that outliers are present in the data.

Figure 2.2 represents the DIF for the three language groups, where the Y-axis represents difficulty and the X-axis the items included in the analysis. Similar ability levels can be observed for the three language groups for most of the items in the vocabulary subtest. However, differences can be identified. The vocabulary items at the beginning of the assessment were very easy for Afrikaans learners. Possibly, the kitchen as represented in the picture is similar to the kitchen in these learners' own homes. Item 10 *'Can you point to the castle?'* was very difficult for Sepedi learners to identify in comparison to English and Afrikaans learners, although these learners also found this item challenging. Item 12 *'Can you point to a pigeon?'* was easier for the Afrikaans learners than the English and Sepedi learners. These items link to exposure

to literature and stimulation in the South African context, whereas in the UK, castles and pigeons are more common. Furthermore, Item 17 *'Can you point to a toadstool?'* was by far more difficult for English learners than for the other two language groups.

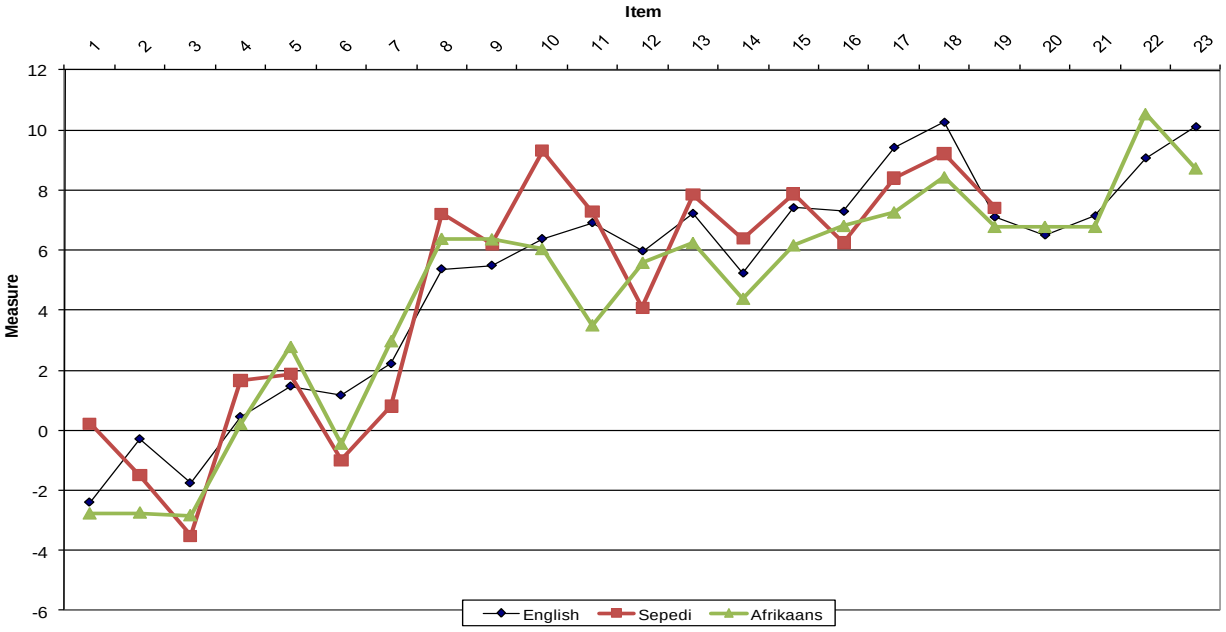

Figure 2.2: Differential item functioning for the Vocabulary subtest for the different languages groups
(Archer, et al., 2010, p. 83)

The DIF analysis indicated that some items functioned differently for the three language groups. These items were examined further to establish if the differences were due to lack of exposure to the stimulus or the translation and graphical challenges experienced throughout the adaptation process.

The statistical examination of the items from the Vocabulary subtest was supplemented by teacher evaluation of the face validity and cultural appropriateness of the items.

Teacher Evaluations

Six Grade 1 teachers (two from each language group) were asked to evaluate the vocabulary subtest by noting if the items and their graphical representations were fair in terms of exposure and culture. They were also asked to rate the difficulty of each item. The results, according to language group for the first two pictures, are seen in Table 2.5.

Table 2.5: Difficulty and values indicated by teachers and assessment of fairness of items

ITEMS	DIFFICULTY			FAIRNESS			
	EASY	AVERAGE	DIFFICULT	CULTURE		EXPOSURE	
				YES	NO	YES	NO
1 - carrots	✪✪ ◎◎ ❖❖	.	.	✪✪ ◎◎ ❖❖	.	✪✪ ◎◎ ❖❖	.
2 - the knife	✪✪ ◎◎ ❖❖	.	.	✪✪ ◎◎ ❖❖	.	◎◎❖❖	.
3 - a fork	✪✪ ◎◎ ❖❖	.	.	✪✪ ◎◎ ❖❖	.	◎◎❖❖	.
4 - a cupboard	✪✪ ◎◎ ❖❖	.	.	◎◎❖❖	.	✪✪ ◎◎ ❖❖	.
5 - some cherries#	◎ ❖	✪✪ ◎◎	❖	◎◎ ❖	❖	✪✪ ◎◎❖	❖
6 - a pan	✪✪ ◎◎ ❖	❖	.	✪✪ ◎◎ ❖❖	.	◎◎ ❖❖	.
7 - a bowl	✪✪ ◎◎ ❖	❖	.	✪✪ ◎◎ ❖❖	.	◎◎ ❖❖	.
8 - the butterfly	✪✪ ◎◎ ❖❖	.	.	✪✪ ◎◎ ❖❖	.	◎◎ ❖	.
9 - the kite	✪✪ ◎◎ ❖❖	.	.	✪✪ ◎◎ ❖❖	.	◎◎ ❖	.
10 - the castle	✪✪ ◎◎ ❖❖	.	.	✪✪ ◎◎ ❖❖	.	◎◎ ❖	.
11 - the wasp	✪✪ ◎◎ ❖❖	.	.	✪✪ ◎◎ ❖❖	.	◎◎ ❖	.
12 - the pigeon#	◎ ❖	✪✪ ◎	❖	✪✪ ◎◎ ❖❖	.	◎◎ ❖	❖
13 - the windmill	✪✪ ◎ ❖	◎	❖	✪✪ ◎◎ ❖❖	.	◎◎ ❖❖	.
14 - the turtle#	✪✪ ◎	◎ ❖	❖	✪✪ ◎	◎❖❖	◎	◎❖❖
15 - the violin	◎◎ ❖❖	.	✪✪	✪✪◎❖❖	.	◎❖❖	.
16 - the padlock#	✪✪ ❖	◎◎	❖	✪✪ ◎ ❖	◎	❖	◎◎❖
17 - the toadstool#	✪✪ ◎❖	◎	❖	✪✪ ◎◎ ❖	.	◎◎ ❖	❖

◎ = English, ✪ = Sepedi and ❖ = Afrikaans

The teacher evaluations of the vocabulary subtest raised issues about the fairness of several items. The cherries item (Item 5) was questioned by one of the Afrikaans teachers, reaffirming the difficulty highlighted by the item analysis. The validity of the graphical representations of Items 12, 14, 16 and 17, (the pigeon, the turtle, the padlock and the toadstool) were also questioned. In some cases of the teacher evaluations, there seemed to have been a tendency to confuse the difficulty of the item with fairness[15]. These cases were discussed with the teacher evaluators to clarify the issue. There was some agreement between the DIF analyses and teacher judgements. The teacher evaluations were supplemented with an expert panel review.

Expert Evaluation Panel

The expert evaluation panel consisted of two research psychologists, two educational psychologists, three teachers, two educational researchers and

[15] The term 'fairness' as opposed to 'validity' was used in the teacher evaluation documents. The document explained that fairness indicates whether the item is appropriate and reasonable to ask, given the children's culture and exposure in their context.

two subject experts involved in teacher education at a tertiary institution. The panel was presented with the above information on the Vocabulary subtest. Particular attention was paid to items flagged by the analyses and teacher evaluations. Concerns were raised by the panel that some aspects in the vocabulary were too Eurocentric, which may have acted as extraneous distracters to South African learners. The possible confounding graphic issues included:

- **Colouring:** Some learners found the colouring alien and distracting. For instance, the colouring used in the kitchen is not commonly found in South African kitchens. This may have confused learners from a poorer socio-economic status.

- **Composition and foreground-background discrimination**: The outside country scene picture was depicted as a view through a window. This caused confusion with many South African learners, as the window did not have burglar bars. Fieldworkers noted that some learners struggled to make sense of the scene, as they did not perceive it as a view through a window. The panel suggested changes to the items, paying particular attention to items flagged in the teacher evaluation.

- **Representation:** This included the windmill depicted in picture two. The windmill depicted was a Dutch windmill. South African windmills look very different with only a metal frame as base. Learners could thus know what the word windmill means, but be unable to link the word to the European representation of the windmill. This would render the item invalid as it would no longer only assess receptive vocabulary, but also exposure to European presentations of the windmill.

Table 2.6: Changes proposed to items by expert evaluation panel

ITEM	PROPOSED CHANGES
4 - a cupboard	May be unfamiliar to learners from very rural areas or with very poor socio-economic status.
5 - some cherries	May consider replacement with a South African fruit (translation into Afrikaans leads to confusion with the candles).
9 - the kite	This is a culturally specific pastime and should possibly be replaced with a more South African item.
10 - the castle	None. This may be a European concept, but learners should have exposure to this through literature. It is also a very well known South African brand name.
11 - the wasp	The drawing of the wasp is inaccurate for the South African wasp species and should be adapted. The translation of wasp into Sepedi is very complex.
12 - the pigeon	The colouring of the pigeon may have to be changed.
13 - the windmill	The item can be maintained by changing the graphic representation of the windmill to the South African windmill.
14 - the turtle	It would be more appropriate to the South African context if this item were changed to tortoise.
15 - the violin	None. Although this item demands a certain level of educational stimulation and exposure, this is true across all language groups.
16 - the padlock	The graphic presentation of this item should be changed. A grey lock with a square shape that is more familiar in the South African context.
17 - the toadstool	Toadstools are relatively unfamiliar in South Africa. “Mushroom” would be an appropriate replacement for this item, with a concurrent change in the graphic representation

Further Examination of Translation and Graphical Elements

Based on the statistical analyses, teacher and expert panel appraisals, the identified items were explored in order to determine how to address the concerns. There was some correspondence between the conceptual as well as empirical processes, with some of the same items being highlighted (items 4, 5, 12 and 17). There were also differences, e.g. the turtle, which functioned well across the groups according to the Rasch analysis and classical test theory, but which was identified as problematic by teachers. It was thus necessary for recommendations and statistical analysis to be weighed against each other in order to decide on the most appropriate adaptations. It was also important to ensure representations were seen as fair and appropriate by teachers to engender trust in the instruments and data.

The first phase was a re-examination of translations. Although a strict protocol of translation and back-translation was followed, there were still some difficulties in translation. Whilst translations were correct, they were sometimes more complex than the original text, thus increasing the difficulties of items in the translated languages.

The translation for Sepedi was very challenging. Since group names instead of specific differentiated words are often employed in Sepedi, some of the translations from Sepedi were academically correct but not often used in the spoken language. Regional dialects of Sepedi are also prolific, leading to complications in the translation process. Careful re-evaluation of the difficulty of the translations was undertaken with a number of translators. In most cases, the translations could be rectified. In the case of the word for 'wasp', no appropriate translation with a similar difficulty value could be identified and a completely new item of a similar difficulty value had to be incorporated into the Sepedi subtest.

After consultation with the CEM, it was determined that some of the more difficult items in the South African test (such as the cherries, saxophone and microscope) were quite advanced in the UK context as well and should not be altered purely because they were more difficult. The problem experienced with the homonym of cherries for Afrikaans learners was addressed by removing the candles on the birthday cake and introducing an extra distracter in the form of a box in the kitchen picture. Alterations to the colouring in the pictures made the items more accessible to Sepedi learners. All graphical elements that had to be changed to be more appropriate for the South African context were first trialled with learners to ensure that the new representations were recognisable to the learners. The adaptation of the instruments resulted in the contextualised SAMP instrument.

2.2.2.3 SAMP 2008

The shift from a computer-based assessment to a paper-based assessment in 2006 proved highly successful. This was true not only for facilitating the adaptation, but also for reducing the operational costs. As a result, the sample of learners in the SAMP project has increased since 2006. The current SAMP

sample is drawn from public schools with Sepedi, Afrikaans or English LOLT in the Tshwane region. Practical constraints such as funding and geographical proximity limited the size and location of the sample. This population therefore represents the accessible population from which the sample was drawn (Best & Kahn, 2006). Multi-phase sampling took place (Cohen, Manion, & Morrison, 2000):

1. Schools where stratified according to LOLT.
2. Eight schools where selected randomly from each LOLT from the DoE databases. A sample of 22 schools was selected on this basis, including two dual LOLT schools.

The sample was inspected at this stage to ensure a fair geographic representation of the Tshwane area and was found to be satisfactory. Currently, some 1,600 learners participate in SAMP each year and are tracked from the baseline to follow-up assessment.

The reliability figures for the SAMP 2008 assessment were well above 0.85 and were comparable to the PIPS UK reliability figures (see Table 2.7). None of the items of the SAMP assessment in any of the LOLT had a negative discrimination value in the 2008 dataset. The inclusive and transparent adaptation process had resulted in an instrument that functions well in the South African context and produced valid and reliable monitoring data that could be used with confidence to inform educational improvement through feedback.

Table 2.7: Reliability coefficients for the three SAMP scales follow-up 2008 and PIPS scales for the UK

SCALE	SAMP 2008	PIPS SCALES UK
Early Phonics	.89	-
Early Reading	.96	.97
Early Mathematics	.95	.90
Total	.97	.98

(Curriculum Evaluation and Management Centre, 2002)

2.3 Conclusion

This chapter provides information on how the PIPS system was successfully contextualised and adapted for use in South Africa for Afrikaans, English and Sepedi. The adaptation process consisted of several aspects, which included the examination of the reliability of subtests and scales; Item facility and discrimination values; Rasch Analysis; Teacher and expert panel evaluations.

The data resulted in several items and aspects of the assessment being flagged for further examination and adaptation. The data from the different processes were weighed against each other and decisions were made for adaptation (see Table 2.8 for a summary).

The reliability values of the newly adapted SAMP instrument indicate that a high quality of data can be generated through the monitoring system. These SAMP data are used in the feedback system that is optimised by this study. Due to the rigorous contextualisation, translation and adaptation process, the data from this system can be used with confidence in Afrikaans, English and Sepedi in the Tshwane region.

Table 2.8: Summary of contextualisation and adaptation of Vocabulary subtest

ITEM	RELIABILITY			ITEM STATISTICS			RASCH ANALYSIS			TEACHER EVALUATION			EXPERT PANEL			ADAPTATION
	English	Afrikaans	Sepedi	English	Afrikaans	Sepedi	English	Afrikaans	Sepedi	English	Afrikaans	Sepedi	English	Afrikaans	Sepedi	
								Kitchen scene								*Colouring altered.*
1- carrots				×	×											✓
2 - the knife					×											✓
3 - a fork					×											✓
4 - a cupboard					×		×	×	×				×	×	×	Colouring altered and 1 door opened slightly.
5 - some cherries					×		×	×	×		×			×	×	Candles removed from cake replacement distracter added.
6 - a pan													×	×	×	✓
7 - a bowl					×		×	×	×							Translation in Afrikaans changed. Translation altered.
								Outdoor scene								*Colouring changed. Burglar bars included.*
8 - the butterfly			×													✓
9 - the kite													×	×	×	✓
10 - the castle							×									✓
11 - the wasp													×	×	×	Graphic presentation and colouring altered. Sepedi item changed to item of equivalent difficulty.
12 - the pigeon								×			×					Graphic presentation and colouring altered.
13 - the windmill													×	×	×	Graphic presentation and colouring altered.
14 - the turtle										×	××					✓
15 - the violin													×	×	×	✓
16 - the padlock										×	×		×	×	×	Graphic presentation and colouring altered.
17 - the toadstool							×				×		×	×	×	Graphic presentation and colouring altered.