

https://dspace.org/

Comparing school based assessments with standardised national assessments in South Africa

Chetty, Mark

2016

http://hdl.handle.net/10539/20787

Chetty, Mark (2016) Comparing school based assessments with standardised national assessments in South Africa, University of Witwatersrand, Johannesburg, <http://wiredspace.wits.ac.za/handle/10539/20787>

Downloaded from a DSpace repository. https://dspace.org/

Comparing school based assessments with standardised national assessments in South Africa

Mark Chetty

Thesis submitted in fulfilment of the requirements for the PhD degree of the University of the Witwatersrand

Johannesburg, 2016

Declaration

I declare that this research report is my own, unaided work. It is being submitted for the PhD degree in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

_____

Signature of Candidate

_____ day of _____ 2016

Abstract

This study compares school based assessments to standardised national assessments in South Africa. The purpose of this research was to understand widespread evidence of a discrepancy in South Africa between the judgements of mathematical achievement made within internal school-based assessments (SBA) and external standardised national assessments (SNA). The study considers the possibility of a mismatch between the design feature rules of SBA and SNA and explores the nature of this discrepancy at a Grade 6 level in three different school settings (suburban, township, and inner-city).

The context of the study was characterised by the aims espoused for assessment in South African national policy interpreted in terms of an activity theory-based nested conceptualization of motives, goals and operational activities. Specific interest was directed at how operations or design features at national policy document level come to function as “rules” for SNA and SBA activity systems. The study was then directed towards explanations and findings on discrepancies between advocated rules and espoused design feature rules interpreted from an analysis of artefacts in SNA and SBA. The key artefacts analysed in the study were the Annual National Assessment (ANA) designed by the Department of Basic Education and the internal SBA assessments designed by Grade 6 teachers.

The study shows that motives and goals that advocate design rules are inferred from a curriculum implementation discourse, with less emphasis on psychological needs and curriculum standards. The enacted activities are considerably varied across school settings. This is evidenced by lower levels of coverage and range and scope of questions in the township school in comparison to the suburban and inner city schools, even when the local district is a protagonist. Across activity systems, comparative analysis of SNA and SBA artefacts showed that there were critical disjunctures leading to deviations from suggested design features, with the state driven ANA and district mediation tools unable to close the gap. The study concludes with notable theoretical and policy implications for thinking about Grade 6 SNA and SBA activity systems in South Africa.

Acknowledgements

Reflecting on my experiences during the course of this study, my sentiments are aptly captured by the expression: “To achieve, is to believe. To attain, is to aspire”.

In this spirit I wish to acknowledge the kind assistance and support of:

My supervisor, Professor Hamsa Venkatakrishnan, who has guided me though the necessary research stages with profound academic rigour and insights that kept me focused, disciplined and enlightened. It is said that “if a task once begun, never leave it till it’s done. Be the labour great or small, do it well or not at all”. Thank you, Hamsa for your continuous commitment, understanding, wisdom and guidance that inspired and motivated me to complete the study.

Professor Willy Mwakapenda for his persistence in getting me started on this journey.

The schools that participated in this research, and especially the teachers, who were willing, eager and passionate participants;

The Marang School of Education for generous financial assistance;

My parents, Sonny and Sheila Chetty, for their positive influence, persistent checking and loving guidance;

My father-in-law Reverend David Naicker and Ms Liziwe Matoti, whose prayers and support brought me to this point;

My family and friends for their understanding and encouragement, and finally;

To my wife Ingrid, and my son Marcel for their continued hope, sacrifice, understanding and inspiration through the process of carrying out and writing up this research study. “From this moment, it is well” – Thank you.

Table of Contents

Index of Figures and Tables vii
1. Introduction 1
1.1 Rationale for focus in this area 1
1.2 Context of the study 7
1.3 The purpose of the study 9
1.4 The problem statement 11
1.5 The research questions 11
1.6 Research design 14
1.7 Background to the study 16
1.8 The structure of chapters 17
2. The literature review 18
2.1 Introduction 18
2.2 Discourses in assessment 20
2.3 The design of SNA 24
2.4 The design of SBA 37
2.5 Categories identified in SNA and SBA for analysing test items 49
2.6 An interpretive map for analysing SNA and SBA 67
2.7 Conclusion 70
3. Theoretical framework 71
3.1 Introduction 71
3.2 An overview of activity theory 72
3.3 Constituting the SNA and SBA activity systems 80
3.4 Looking within and across activity systems: Overlaps, tensions and contradictions 84
3.5 Conclusion: Applying an AT framework to this study 88
4. The Research Design 92
4.1 The research design 92
4.2 Sampling 94
4.3 Data collection 96
4.4. Document analysis 98
4.5 The interview process 106
4.7 Ethical considerations 110
4.8 Conclusion 111
5. Standardised National Assessment in South Africa 112
5.1 Introduction 112

5.2 The hierarchical structure of design rules in SA-SNA 113
5.3 EMT analysis on the design rules 123
5.4 The assessment artefacts 131
5.5 Conclusion 151
6. School based assessment 154
6.1 Introduction 154
6.2 The advocated design rules of SBA 155
6.3 SBA tool formats 166
6.4 Assessment activity within an EMT framework 174
6.5 Considering the mathematics seen in assessment artefacts 190
6.6 Conclusion 246
7. Summary, key findings and conclusion 249
7.1 Introduction 249
7.2 Summary of key findings within SNA and SBA 249
7.3 Looking across SNA and SBA: discussion of findings and implications 251
7.4 Concluding remarks 264

References 268

Annexures 281

Index of Figures and Tables

Table 1.1	Ratings Scale for learner achievement	2
Figure 1.1	Grade 3 Comparative numeracy scores for 2001 and 2007	3
Figure 1.2	Grade 6 mathematics achievement levels per province in South Africa	4
Figure 1.3	Grade 6 mathematics results of the 2008 ANA	8
Table 2.1	Summary of Morgan's mainstream assessment discourses	23
Table 2.2	Examples of SNA in mathematics in South Africa	29
Table 2.3	Target Percentages of the TIMSS 2011 Mathematics Assessment	31
Table 2.4	SACMEQ II scores for Grade 6 mathematics by quintile and country	33
Figure 2.1	Specimen of a "test framework" from the Grade 6 SE study	36
Table 2.5	A mathematics two-way chart on Fractions relating instructional objectives to instructional content	42
Table 2.6	A mathematics table of specifications on Fractions and Decimals	43
Figure 2.2	Anderson's revised Taxonomy of Educational Objectives	44
Table 2.7	The LTA Taxonomy	45
Figure 2.3	de Lange's assessment pyramid	47
Table 2.8	Number of ASs per LO in the RNCS	52
Table 2.9	Mathematics skill levels from the Laos Assessment survey	58
Table 2.10	Webb's DOK classification system	60
Table 2.11	Example of a Multiple-Choice Item	65
Table 2.12	Example of a Closed Constructed-Response Item	66
Table 2.13	Example of an Extended Response Item	66
Table 2.14	A literature map for analysing test item categories in SNA and SBA	69
Figure 3.1	First generation activity theory: Vygotsky's (1978) model of mediated action.	74
Figure 3.2	Leont'ev's hierarchical structure of activity.	75

Figure 3.3	Engeström's (1987) expanded mediation triangle (EMT)	76
Figure 3.4	Third generation AT model of interacting activity systems	79
Figure 3.5	Hierarchical structure of activity	85
Figure 3.6	Interacting assessment activity systems	90
Figure 4.1	Triangulation of data	97
Table 4.1	Themes, design features and sub-categories	101
Table 4.2	Tool A: Mathematical coverage in an assessment task	102
Table 4.3	Selection of common topics	104
Table 4.4	Item characteristics for the selection of mathematics	105
Table 4.5	Exemplar item map for SNA	106
Table 5.1	The five different types of assessments listed in the RNCS	113
Table 5.2	Principles of OBA	114
Table 5.3	Extract of the ANA 2010 Grade 6 Mathematics test framework	121
Table 5.4	The difficulty levels	122
Table 5.5	The cognitive levels	122
Figure 5.1	EMT model for the SA–SNA Activity System	124
Table 5.6:	SA-SNA goals located within different assessment discourses	125
Table 5.7	Aggregated counts of LOs	132
Figure 5.2	LO % distribution 2008–2010	133
Figure 5.3	Coverage of assessment standards in the ANA tests	134
Table 5.8	Difficulty level descriptors	136
Figure 5.4	Spread of DL % in the ANA tests from 2008–2010	138
Table 5.9	CD descriptors	140
Figure 5.5	CD spread for ANA 2008–2010	143
Table 5.10	LD descriptors	145
Figure 5.6	LD % spread across for the ANA tests: 2008–2010	147
Table 5.11	IF descriptors	150

Figure 5.7	IF % spread for the ANA tests: 2008–2010	151
Table 6.1	Pass rate for Mary's Grade 6 classes in term 1 in 2010	161
Figure 6.1	IC-school Grade 6 mathematics work schedule	163
Table 6.2	Assessment tasks for Kalay	167
Table 6.3	Assessment tasks for Fiona	169
Table 6.4	Assessment tasks for Mary	171
Table 6.5	Kalay's SBA Activity system	177
Table 6.6	Fiona's SBA activity system	184
Table 6.7	Mary's SBA activity system	188
Table 6.8	The mathematics selected in Kalay's overall SBA	192
Table 6.9	Kalay's AS selections within each Task	194
Table 6.10	The mathematics selected in Fiona's overall SBA	195
Table 6.11	Fiona's AS selections within each Task	196
Table 6.12	The mathematics selected in Mary's overall SBA	196
Table 6.13	Mary's AS selections within each Task	198
Figure 6.10	Percentage distribution of LO for the overall SBA	200
Figure 6.11	Comparative weightings for Mary	201
Figure 6.12	Coverage of non-repetitive AS in SBA	203
Figure 6.14	LO coverage – Year-Long SBA vs Year-end examination	205
Table 6.14	Exemplar item map for Fiona	206
Table 6.15	Examples on difficulty level within LO 1 – Equivalent Fractions	208
Figure 6.16	Difficulty level - Examination	209
Table 6.16	Examples of CD coding	210
Figure 6.17	Cognitive demand (%) - Examination	212
Table 6.17	Examples of LD coding	213
Figure 6.18	Language demand (%) - Examination	215
Table 6.18	Examples of IF coding	216

Figure 6.19	Item format (%) - Examination	217
Table 6.19	Summary of the range and scope of questions	218
Table 6.20	Common assessment topics according to LO	219
Table 6.21	Assessment tasks on Ratio	222
Table 6.22	Summary of teachers' assessment of Ratio	227
Table 6.23	Assessment tasks on Capacity	231
Table 6.24	Summary of teachers' assessment of Capacity	235
Table 6.25	Assessment tasks on Graphs	238
Table 6.26	Summary of teachers' assessment on Graphs	245
Figure 7.1	Elements of interacting policy layering	252
Figure 7.2	Parallel spaces across interacting activity systems	255
Figure 7.3	Mathematics coverage: SNA and SBA	257
Figure 7.4	Cognitive demand comparison - ANA vs School Examination	258
Figure 7.5	Language demand comparison - ANA vs School Examination	259
Table 7.1	Exemplification of item characteristics	260
Table 7.2	Design features for establishing the range and scope of questions	262

1. Introduction

This study is focused on widespread evidence of a discrepancy in South Africa between the judgements of mathematical achievement made within the interacting assessment systems of internal school-based assessments (SBA) and external standardised national assessments (SNA). I explore the nature of this discrepancy at Grade 6 level and offer an analysis. The context of the study was linked to the introduction of standardised Annual National Assessments (ANA) implemented within the Foundations for Learning (FFL) campaign launched by the Department of Education (DoE) in 2008 in all public schools in South Africa. Activity theory is utilised as a theoretical lens to explain the discrepancy and a case study method is used for generating evidence. The study is detailed below.

1.1 Rationale for focus in this area

In 2008, the constitutionally agreed outcomes based education (OBE) approach was applied from entry level (Grade 1) in the foundation phase to the exit level (Grade 12) in the Further Education and Training (FET) phase for the first time, following waves of reform in the post-apartheid years which first focused on the General Education and Training (GET) phase through the Revised National Curriculum Statement (RNCS) introduced in 2002, and then shifted attention to the FET phase curriculum reform through the national curriculum statement (NCS) introduced for the first time in Grade 10 in 2006 (DBE, 2009). With the phasing-in of the revised curriculum complete, state institutions began focusing increasingly on improving the quality of curriculum delivery and learner attainment, with assessment flagged as a high priority within this focus (DBE, 2009).

Recent studies in South Africa have indicated that whilst revised curricula have been put in place, and a degree of resources and training has been provided, learners continue to perform below the standards of the prescribed curriculum (Kanjee, 2006) and well below the performance levels identified in internal school assessments (van der Berg and Louw, 2006) when tested on standardised assessments. Additionally, there have been notable discrepancies in learner performance between internal portfolio assessments and levels of performance in the final senior certificate mathematics examinations with the final results significantly lower than levels

allocated in teachers' assessment of portfolios (van der Berg & Louw, 2006). Given the low scores obtained by South African learners in international assessments like the Trends in International Mathematics and Science Study (TIMSS) and national assessments like the Systemic Evaluation (SE) study, and given the difficulties teachers are described as having in assessing learners at the primary school level by the Department of Basic Education (DBE, 2009) [1] it appeared reasonable to expect that similar discrepancies existed in primary schools as well.

While the National Protocol on Assessment (Grades R-12) standardised the recording and reporting processes for schools (DoE, 2005d), the lack of standardisation of test criteria means a 'pass' can relate to different things in different schools. For the Intermediate phase (Grades 4-6) a 4-level ratings scale (**Table 1.1**) was used to determine the level of learner achievement for internal tests and grade progression. Passing in terms of this ratings scale loosely translated into learners achieving a level 2 or higher.

Table 1.1: Ratings Scale for learner achievement

Level	Descriptor	%
1	Not Achieved	1-34
2	Partially Achieved	35-49
3	Achieved	50-69
4	Outstanding	70-100

(Source: DoE, 2006: National Protocol on Assessment)

At the primary school level, alarmingly low learner achievement scores have been recorded in national and international standardised assessments (ANA, 2012; PIRLS, 2006; SE, 2004; TIMSS, 2003; SACMEQ, 2000). In 2012 the ANA results revealed that poor performance of learners continues to be a concern for the education sector, where participating Grade 6 learners in public schools scored an average of 26.7% in Mathematics. This average has risen up to 43.1% in 2014. Recent studies have pointed out reasons for the underperformance of learners in SNA in South Africa (Reeves & Muller, 2005; Spaull, 2008; Taylor, 2008). van der Berg and Louw (2006) described the problem as follows:

Many Grade 6 learners are not able to perform mathematics tasks expected at

[1] Prior to 2009, the Department of Basic Education (DBE) was referred to as the Department of Education (DoE).

the Grade 3 level.

A further problem is that there is a bimodal pattern of performance, with performance varying significantly between high socio-economic status and low socio-economic status schools (Spaull, 2008).

The picture is somewhat different when considering internal pass rates on the SBA at school level. Between 1999 (where the pass rate was 48.8%) and 2009, there has been an average annual increase in the pass rate in primary education (i.e. those who passed expressed as a proportion of those who wrote) of six percentage points (DoE, 2003). Since the National Protocol on Assessment in South Africa had given teachers the 4-level score (**Table 1.1**) against which to assess learner performance on the assessment standards covered in any internal assessment in 2005, promotion rates pointed towards 80%-95% of learners progressing to the next grade each year. Internal assessment data points to the vast majority of these learners recorded as passing mathematics in the previous grade with high levels of achievement contradicting the findings of the SNA studies conducted in South Africa.

The 2008 Grade 3 SE study results indicated that the average score for Numeracy was 35% with only 15% of learners in South Africa able to pass both Numeracy and Literacy tests (DoE, 2008b).[2] The results showed only a 5% improvement in Grade 3 numeracy results between 2001 and 2007. This showed the urgent need to improve performance in critical foundational skills (DoE, 2008b) across all nine provinces.

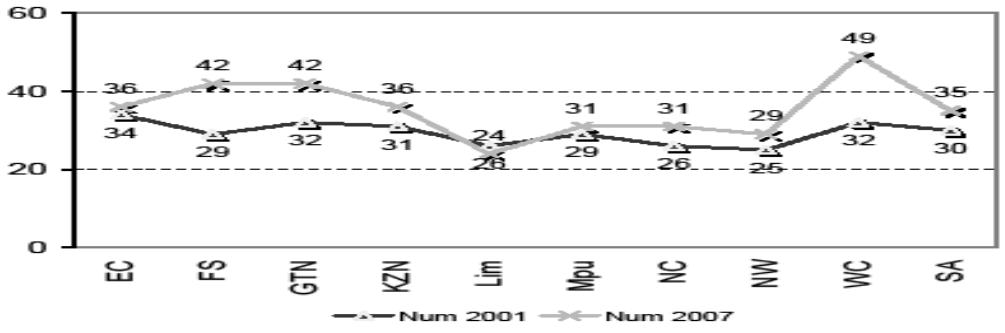


Figure 1.1: Grade 3 Comparative numeracy scores for 2001 and 2007

(Source: DoE, Systemic Evaluation Grade 3 Results)

At the Grade 6 level, the SE study conducted by the DoE in 2004 reported low

[2] This Systemic Evaluation (SE) survey was conducted in 2007 with a random national sample of 54 449 grade 3 learners from 2 355 primary schools.

achievement levels in mathematics with 81% of learners considered non achievers (DoE, 2005). As with the Grade 3 study, the underachievement in Grade 6 was widespread across all nine provinces (see **Figure 1.2**). The mean score across learner achievement in mathematics was 27% (DoE, 2005a).

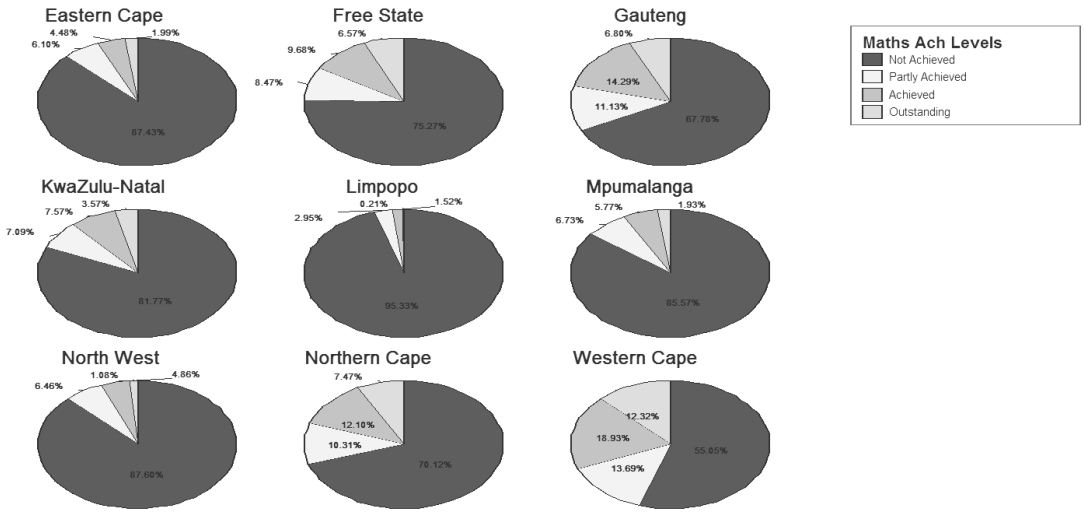


Figure 1.2: Grade 6 mathematics achievement levels per province in South Africa

(Source: DoE, Grade 6 Systemic Evaluation Report, 2005)

Considering the contrasts between the SBA and SNA scores obtained by South African learners in mathematics and languages, local researchers were increasingly suggestive of discrepancies existing between internal SBA and SNA in mathematics at a primary school level (Soudien, 2008; Kanjee, 2006; Chisholm, 2005; Taylor & Vinjevold, 1999). There are a range of reasons for these discrepancies with the possibility of varying goals at play, namely, school and policy level pressure to meet targets for learner attainment (e.g. national targets calling for increasing learner performance in mathematics from 30% to 50%), a desire not to hold learners back and teachers delivering the curriculum at minimal levels of achievement.

A further possibility, given the evidence of discrepancies in test results presented earlier, is that there could be mismatches between the design of SBA and the design of SNA. Although structured by common national curriculum policy prescripts the design features of SBA and SNA could be different in their underlying frameworks or "rules", and in the case of SBA further underlying differences could be revealed if the frameworks are structured differently as a result of varied teacher interpretations

across school settings. Unpacking the rules advocated for the design of SNA and SBA assessments thus became the initial thrust of the study.

Through my professional role in the DoE National Assessment unit, I had observed that key features underlying the design of SNA including: content coverage, range of levels of difficulty, range of cognitive and language demands of items, different question formats and representations were frequently absent or limited in teachers' design of the SBA used to make judgements about learner performance. Later in this study I detail and critically discuss the frameworks (rules) against which the standardised tests (in particular the ANA) and the SBA analysed in this study were based.

Evidence in South Africa showed that teachers applying the RNCS often lacked the required content knowledge to design appropriate assessments (Chisholm, 2005; Taylor & Vinjevold, 1999). It has been suggested that teachers have difficulty interpreting aspects of the curriculum because of under specification in the RNCS in both content and progression which limits openings to effectively guide teachers to 'design down' assessments from the outcomes (Brodie, 2008; Taylor & Vinjevold, 1999). Broader support for this under-specification has also been presented in Hoadley and Jansen (2009) and Reeves and Muller (2005) who have all argued for greater specificity in the curriculum. Hoadley and Jansen (2009) found that there was a gap between the curriculum plan and the assessment practices of teachers. Understanding how teachers designed their assessments against what was advocated became the second thrust of the study.

The collective argument made from these findings was that under-specification of the RNCS made it difficult for teachers to develop and benchmark internal SBA in mathematics against an expected standard and that this difficulty was exacerbated for teachers whose content knowledge was weak and in contexts with limited access to useful resources. The specifications teachers received on the RNCS and policy directives on Outcomes Based Assessment (OBA) for school-based assessments were mainly in the form of generic principles (DoE, 2002b). OBA principles generically explained assessment practice but were not learning area specific to mathematics.

A ministerial review study conducted by the DoE on the evaluation of the NCS in the intermediate phase found that teachers were not sufficiently competent on applying OBA principles within their learning areas and that forms of assessment were mainly driven by summative goals (DBE, 2009). The study indicated that assessment was mainly viewed as tasks for completion of portfolios instead of a planned series of different events, activities and actions to gather as much information as possible about learners' progress (DBE, 2009). One of the key findings of the study was that the request by teachers to be further assisted on assessment in the RNCS needed to be taken seriously.

Although much research has been conducted in establishing learner achievement scores in South African public schools using standardised national and international assessments (ANA, 2012, 2011, 2009, 2008; PIRLS, 2006, 2011; SE, 2005, 2007; TIMSS, 2003, 2011; SACMEQ, 2000, 2007) little is known about the nature of the discrepancies that exist between internal SBA and SNA in mathematics, that may feature within the broad low level performance. At primary school level in particular, there is a dearth of empirical evidence on the matter.

At this stage research is certainly lacking on the nature and specificity of internal SBA at the primary school level in mathematics, and what aspects figure within teachers' selection of mathematical content and range of tasks to include in their assessments. This study adds clarity on overlaps and contrasts between policy advocated rules governing assessment task design in SNA and SBA and the ways in which the assessment tasks in both assessment systems reflect what was advocated.

The main goal of classroom testing and assessment is to obtain valid, reliable and useful information concerning learner achievement (Linn and Miller, 2004). The low performance of learners in national and international assessments, while simultaneously being judged as making adequate progress in internal school testing, motivated the rationale for this research: a need to understand at a primary school level, the nature and specificity of differences that exist between internal SBA and the externally driven SNA in mathematics.

1.2 Context of the study

The context of the study was characterised by the aims espoused for assessment in South African national policy interpreted in terms of an activity theory-based nested conceptualization of motives, goals and operational activities. While these terms are defined and explained later in Chapter 3, it is enough to note at present that in this study, motives were regarded as broad collective aims of the education system, goals were directives specifically related to assessment within this system, and operational activities were operations executed in the name of motives and goals. Further theoretical detail on my use and understanding of these terms follows in Chapter 3. Within an assessment landscape that had direct policy linkages to revised curriculum statements, improvement plans and interventions and the introduction of national assessments, the discourses relating to SNA and SBA were considered for further investigation.

Recent policy debates in South Africa on moving education forward two decades after the democratic transition (in 1994) have been frequently released in media reports (Ramphele, 2009) and political discussion documents (Education Roadmap – Focus on schooling system, 2008). A priority in the 10-point improvement plan listed in the 2008 Education Roadmap was the sustained focus on improving Numeracy and Literacy scores of learners (DBSA, 2008). This focus was aligned to the improvement plan articulated in the FFL campaign launched by the DoE in 2008. The FFL campaign was introduced with the following key features:

- Breakdown of the curriculum into milestones[3] (knowledge and skills).
- Support teachers to plan assessments for the year.
- Assist teachers to monitor learner progress.
- Assist teachers to develop the required assessment tasks per term.
- Introduce annual national assessments (ANA) (DoE, 2008a).

With the launch of the FFL Campaign, the broad goals of the Education Department became more directed towards supporting teachers to deliver the expected

[3] The term 'milestones' is used to indicate the expected level of development of learners' progress to becoming literate and numerate. It gives a sense of what their achievements could be at given points in the school year across the primary grades 1 to 6 (DoE, 2008).

curriculum with progression, to assist teachers in planning assessments aligned with the curriculum and to measure performance through the use of standardised assessments. Standardisation in both teaching (in terms of content coverage, sequencing and pacing) and assessment practice became an explicit policy goal. Targets were set for measuring the success of the FFL campaign in relation to improvement in learner performance.

> *'The projected measure of the campaign is to increase average learner performance in languages and mathematics to 50% in 2011 and learner assessment will occur on a regular basis with standardised assessment tasks provided by the DoE' (DBSA, 2008).*

There was an intention to improve transparency and accountability in assessment in order to give teachers and learners a clearer understanding of the expectations for internal assessments and the knowledge, skill and values to assess. The policy focus was to provide standardised assessments and a milestone curriculum to support teachers to design better internal assessments. This included exemplification of the curriculum with a high degree of prescription of skills and assessment coverage for each term with examples of assessment tasks and rubrics within and across primary grades. To this end, in 2008, the DoE conducted the first batch of annual national assessments (ANA) as part of the FFL campaign targeting primary grades 1-6 in the General Education and Training Band (GET) band.

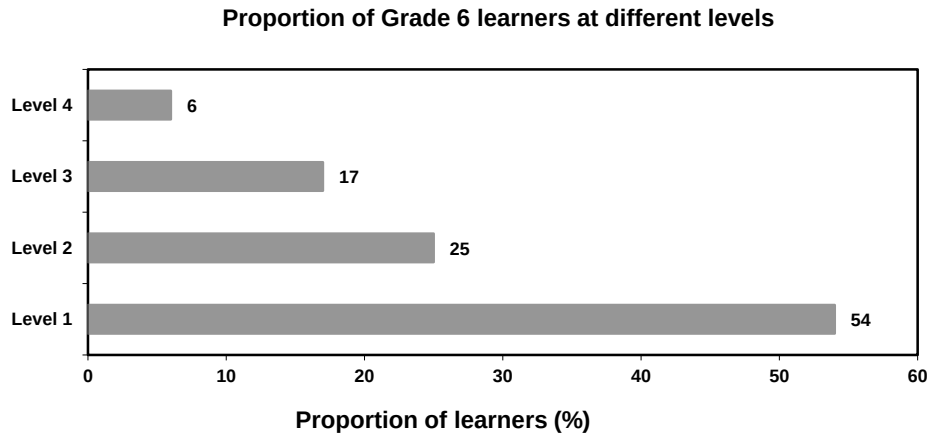


Figure 1.3: Grade 6 mathematics results of the 2008 ANA

Findings from the first round of the ANA confirmed the trend of low learner performance in standardised tests in mathematics. In the Grade 6 results for

mathematics (**Figure 1.3**), 54% of learners did not achieve the required assessment standards (level 1) with only 6% of learners placed in the outstanding group (level 4). These trends in performance were mirrored in the provincial results and, as indicated earlier, continue in recent results. Although the initial focus of the ANA was on improving learner performance, its introduction in the assessment landscape of South Africa also put in place an accountability mechanism for curriculum implementation that aimed to align internal teacher assessments with State driven SNA.

> *ANA should encourage teachers to assess learners using appropriate standards (DBE, 2011, p.10)*

Within these broad motives of policy pressure by the state to strengthen school-based assessments and promote greater alignment between SNA and SBA, this study's focus was on exploring the operational activities of internal SBA in three schools drawn from varied socio-economic and historic privilege strata. At the level of operations, the introduction of the ANA as part of the FFL campaign provided a relevant context and a suitable assessment artefact in which to try and understand in depth, possible discrepancies between internal and external assessments of mathematics and for such an investigation to be conducted within a specific grade (Grade 6) in a particular province (Gauteng) with a small sample of mathematics teachers from different school settings.

In this study, three different school settings (suburban, township, and inner-city) were considered. An important focus of the study was to provide in-depth comparisons of SBA with SNA in different school settings, and to explore reasons for any discrepancies found in relation to school contexts. The schools were selected purposively on the basis of explicit historical reporting of discrepancies in learner performance between internal and external assessments, and thus, were representative of this broader research finding in the South African landscape.

1.3 The purpose of the study

The purpose of the study is to better understand disparities between the advocated assessment policy discourse at a national level and the enacted practices of designing assessment artefacts for SNA and SBA. The intention was to generate and analyse evidence for an improved understanding of assessment of mathematics

at a Grade-6 level. It was therefore useful to look at how policy related motives and goals advocated for assessment at the national level influenced the operational activities (rules) of designing assessment tasks. The analysis of operations focused on design features related to curriculum coverage and the range and scope of questions, all features that figure in the literature base on SNA and SBA. Specific interest was directed at how operations or design features at national policy document level come to function as "rules" for SNA and SBA. This study was directed towards explanations and findings on discrepancies between advocated rules and espoused design feature rules interpreted from an analysis of artefacts in SNA and SBA and then to look across them.

As noted earlier, primary schools operate within policy goals related to throughput and promotion (DoE, 2003b), as well as those related to the improvement of learner performance in literacy and numeracy (DoE, 2008b). In terms of the data gathered, and by looking at *teachers' reasons* for designing or selecting specific tasks, the hope was to understand their interpretations of the mathematics curriculum, the basis of their judgements on learner capabilities and their understanding of wider school and systemic goals that they considered relevant within their justifications for the tasks they used.

By *analysing tasks* educators used to judge performance in mathematics in SBA and comparing them to those used in the SNA, discrepancies and contradictions were highlighted within and between the two types of assessments. SBA and SNA share and are shaped by a common goal: the need to assess learners' capabilities in mathematics. As noted earlier, recent studies and reports have showed that disparities may have existed because teachers found it problematic to assess learning against an underspecified curriculum (DBE, 2009; Brodie, 2008; Chisholm, 2000; Taylor & Vinjevold, 1999). By looking critically at design rules and their *application in assessment artefacts* of SNA and SBA, this study investigated the evidence gathered to explain disparities between these two assessment contexts.

With explicit goals attached to the introduction of annual national assessments in mathematics, an opportunity existed for research to be conducted that looked specifically at the motives, goals and operational activities guiding teachers' design

of SBA and making comparisons to the policy related motives, goals and operational activities specified for SNA. Comparative analyses of selected tasks against advocated rules were used to describe empirical data on discrepancies within and across SNA and SBA.

1.4 The problem statement

This research was problematised in the following statements: National policy documentation in South Africa advocates “rules” for the design of SNA and SBA, but evidence suggests that these seem not to be clearly applied. To explain the nature of the problem, the discourse relating to SNA and SBA were explored analytically on two levels: 1) the advocated design features or rules for designing SNA and SBA with a key focus on their underlying motives and goals, drawn from the literature, and 2) the enacted design features interpreted from an analysis of test items considered in SNA and SBA assessment artefacts.

The underlying theoretical view of the research was to look at SBA and SNA as two distinct, yet complementary and interacting, activity systems. In defining an activity system, a deliberate attempt was made to unpack essential elements that constituted such systems, as proposed by the theoretical viewpoints of activity theory (AT) proponents such as the Soviet psychologist A.N Leont’ev and the Finnish researcher Yrjo Engeström. Using AT concepts, descriptive commentary was provided on the components of SBA and SNA and the nature of their existence as activity systems of assessment. Specific attention in the narrative was given to discussing and comparing critical disjunctures within and across SNA and SBA activity systems. In Chapter 2, a literature map for analysing enacted design features is presented. This literature based on advocated design rules provided a vantage point for considering both the design rules advocated within the SNA and SBA systems, and the enacted rules seen within their respective assessment artefacts.

1.5 The research questions

Within the SNA and SBA contexts, design rules influenced by the broader motives, goals and rules found within each system were explored and analysed according to three research questions. Specific categories drawn from the literature on assessment generally and in mathematics education specifically allowed for the

development of an analytical framework to explore the detail in the data generated on the advocated assessment rules and the enacted rules found in the assessment artefacts. It was on the basis of specified categories and sub-categories that this research study looked to explore and understand rules in SNA and SBA.

The following research questions formed the focus of the study:

1. What is the nature of design rules for SNA and SBA in Grade 6 Mathematics for assessing learners?
    a. What motives and goals for assessment (viewed as advocated rules for the design of assessment artefacts) can be identified within South African policy documents for SNA and SBA? What overlaps and tensions exist between the motives and goals within these advocated rules?
    b. In what ways do the SNA artefacts and the SBA artefacts in three schools in different socio-economic contexts reflect the rules advocated? What overlaps and tensions are seen within the enacted rules found in the assessment artefacts and those that were advocated in the broader motives and goals?
2. What variations exist in the assessment artefacts among teachers in different school settings?
3. How do the internal SBA in the different schools compare with external SNA for assessing learners in Grade-6 Mathematics?

The first question was divided into two sub-questions. The first part focusses on the overall rules advocated for SNA and SBA. Given the South African policy context and its explicit aim of linking SNA and SBA activities, it was important to understand the rules underlying SNA and SBA both separately and as complimentary activity systems in order to judge the extent of overlap. Structurally, I had to first explain the rules of SNA and SBA separately and then explain the extent to which SNA rules figured in the SBA context. Through this analysis, as noted already, I remained alert to the assessment research base, and what findings drawn from this base alerted me to in terms of affordances, tensions and constraints.

For the second part, the advocated rules were analysed according to test item

categories drawn from mathematics education and assessment research. Categories used in international studies (e.g. TIMSS) and by the DoE in their national SE studies and the ANA pointed towards mathematical coverage, difficulty level, cognitive demand and item (question) format as aspects to consider as design rules. Within each of these design features, empirical evidence was generated and analysed against pre-defined categories and sub-categories developed with reference to the policy context and the literature to interpret findings. In this study, I critically reflect on this range of categories and sub-categories and use them as a common frame to analyse both SBA and SNA.

The emerging evidence from prior research has suggested that the design of assessment tasks is seen as a measure of quality teaching and the most important predictor of student achievement (Darling-Hammond, 2000). The focus of the research questions was however not on student achievement and quite often a distinction had to be made in the data generated between the analysis of assessment tasks and analysis of learner performance. The latter figured only in the context of teachers' descriptions of overarching goals and purpose of assessment.

At primary school level, there is generally insufficient research information available on the actual assessment artefacts teachers' use, and the obstacles encountered in their design of assessments are often not clearly revealed for corrective action to be taken (van der Berg, 2005). Prior evidence has suggested that one such obstacle relates to the teachers' design of internal tests with an inappropriate match to desired learning outcomes and assessment standards. In the second part of the first research question the nature of the rules advocated for assessment design in SNA and SBA were investigated.

The second research question was a follow up activity to the first question. Here, the teachers' assessment tasks are investigated in different school settings with varying socio economic status. This question was relevant in a polarized educational context that is highly resource driven (Reeves & Muller, 2005), that affects teachers' design and use of assessment tasks. Research from prior assessment studies such as SACMEQ (Moloi & Strauss, 2005) and the TIMSS (Reddy, 2006) has indicated that South Africa has unusually high between-school differences as opposed to the norm

of high in-school differences observed in other participating countries (van der Berg & Louw, 2006). This context motivated the second research question, where variations in teachers' design of tests in schools from different settings were investigated.

The third question dealt with comparing the two assessment systems (SBA and SNA) against the sub-components (a. and b.) used in the first research question. The focus here was to look across SNA and SBA by first looking at an AT analysis taking in motives and goals that underlie the relationship between advocated design rules and enacted tools; and then to look at a mathematical analysis of tools by comparing coverage and range and scope of questions with what was advocated Within this study elements for comparing SNA to SBA, drawn from the literature are presented and discussed. I also conclude the study with measures for analysing SBA and SNA using a common framework of categories that provide a suitable basis for test design and comparison.

1.6 Research design

As this study was exploratory and interpretive, a qualitative research approach was followed. Researchers working within the qualitative research paradigm investigate the quality of the relationships (e.g. between teachers and materials) and try to understand activities (e.g. assessment practices) more fully (Fraenkel & Wallen, 1990). Qualitative researchers usually work in an exploratory-descriptive way and use sensitizing and empathizing rather than quantifiable concepts (Booyse, 1993). This is not to say that the study operated exclusively without any consideration to the use of quantitative or mixed methods to illuminate points of interest that arose. In the analysis chapters, selected data sets on SNA and SBA are aggregated and presented in the form of graphs.

Data collection and analysis were, to a large extent, determined by the nature of assessment activities evidenced in the artefacts of SNA and SBA. Empirical evidence for SNA in this study was grounded in both curriculum artefacts such as the RNCS, FFL and the education sector plan and in assessment artefacts consisting of ANA papers from 2008 to 2010 supplied by the DBE. Empirical evidence for SBA could be found in the RNCS, FFL, policy guidelines and in the multiple internal SBA

tasks designed by the Grade 6 teachers responsible for mathematics teaching in three different school settings over an entire academic year. Teacher records included: national policy documents and guidelines, local district material, formal internal assessments used for learner progression, assessment planning records, learner assessment records, and school assessment policies that informed practice.

With the introduction of the FFL, ANA was introduced to all public schools having Grades 1-6 (DBE, 2008a). There were now multiple data sources in SNA and in mathematics teacher's tasks (SBA) available for viable research to be conducted. The ANA Grade 6 Mathematics papers were the specific SNA artefacts considered, while the other policy documents were regarded as source documents providing input on rules and/or motives and goals. Teachers' assessments were looked at both as individual constructs and as artefacts of a collective Grade 6 Mathematics community constrained by individual teacher knowledge and their access to resources.

Schools were chosen from three different education districts within the Gauteng province so that evidence could be generated on teachers' assessment tools that emanated from different approaches but set within the constraints of singular provincial formats and national guidelines. Differences in assessment strategies due to mediation of different curriculum support and intervention strategies at the local district office level were taken into account within the analysis. Additionally, I documented variations in the design of SBA in a school from a township (with low socio-economic status), an inner-city (with middle socio-economic status) and a suburban (middle to high socio-economic status) area.

Merriam (1998) argues that 'cases' are studied in their own right, not just as samples from populations. In this way, the generalization of findings to a broader population or community was not a primary issue of concern in case study designs. Fraenkel and Wallen (1990, p.370) stated that much can be learned from 'studying just one individual, one classroom, one school or one district.' As this study was about investigating the nature of the differences between SNA and SBA, it was important to select schools and teachers that were 'information rich' in their assessment activities. My purposive selection of schools was based on the following evidence:

- availability and access to the assessment tasks and records of Grade 6 teachers,
- participating in and implementing the annual national assessments,
- school management supportive of the study and willing Grade 6 teachers,
- socio-economic status of the school, and,
- location and proximity of schools as viable research sites.

In a case study design, it is important to use data collection methods that allow the researcher to identify and interpret the phenomenon being investigated. In this study document analysis and interviews were the preferred data collection methods to investigate the nature of assessment activities. Although I was mainly interested in documented data on assessments, looking at learner scripts provided a route into triangulating teacher claims on tasks used in SBA.

In the write up of the analysis chapters, I provide a critical analysis on SNA and SBA as interacting activity systems with interpretations and commentary made on structure, mediation, tensions and contradictions, culminating in an examination of comparisons within and across them.

1.7 Background to the study

> *Our current assessment systems are harming huge numbers of students for reasons that few understand. And that harm arises directly from our failure to balance our use of standardized tests and classroom assessments in the service of school improvement. When it comes to assessment, we have been trying to find answers to the wrong questions (Stiggins, 2002).*

This study is an attempt to find the right answers to the right questions. It investigated the nature of advocated design rules and compared them to those enacted in actual assessment artefacts. This area of research is regarded as a gap that needs to be urgently addressed in the South African context. This study is therefore aimed at making a meaningful contribution towards a better understanding of an assessment landscape that continues to be marked by substantial inequity.

1.8 The structure of chapters

Subsequent chapters of the study are structured as follows. In Chapter two, I present a literature review of international writing on SBA and SNA concepts. In Chapter three, I discuss the theoretical underpinnings of the study, specifically drawing attention to the use of AT and how its applications featured in the research methodology and design. In Chapter four, I discuss the research methodology of the study linking empirical evidence to the data sources used. In Chapter five the SNA is discussed with the standardised ANA considered as the key SNA artefact. In this chapter, I offer a critique of the South African SNA activity system during three years of ANA implementation from 2008 to 2010, using analytical categories constructed as reference points for discussion of design rules relating to coverage and the range and scope of questions. In Chapter six, I present an analysis of empirical data relating to internal SBA activities within concepts associated with AT loci of the rules, goals and tools of assessment based on constructed themes similar to that used in Chapter 5. Empirical data is linked directly to the research questions. In Chapter seven, I conclude by comparing the SNA and SBA findings through juxtaposing the empirical evidence from the two previous chapters and relate this back to the literature consulted. Key findings are illuminated and summarised. A list of acronyms as well as definitions of terms has been included as annexures to enhance the readability of the thesis.

2. The literature review

2.1 Introduction

A central feature of this study is to explore two types of assessment systems: SNA and SBA. International literature indicates that SNA refers essentially to national or cross-national assessments that aim to provide systemic data about the achievement of learning outcomes defined by the State or agency (either local or international) implementing them. Systemic data refers to how learners are performing as a collective rather than as individuals. SNA is often accompanied by policy-level targets at the aggregate level, for example, a systemic target outcome could be that all schools should achieve at least 50% as a mean score in mathematics in Grade 6. The Dakar Framework for Action in 2000 stressed the importance of having “a clear definition and accurate assessment of learning outcomes (including knowledge, skills, attitudes, and values)” as governments worked towards ensuring the provision of quality education for all (Postlethwaite and Kellaghan, 2008).

SBA, in contrast, consists of assessment embedded in the teaching and learning process (Braun & Kanjee, 2006). These assessments are usually devised and administered by class teachers, although some are the work of the school principal or other teaching staff. SBA has a number of important characteristics that distinguish it from SNA. Generally, SBA is aligned with the delivered curriculum and may employ a broader array of media (e.g. oral presentations) and address a greater range of topics than is the case with SNA (Braun & Kanjee, 2006). Davison and Hamp-Lyons (2009) identified the following key characteristics of SBA:

- It involves the teacher from the beginning to the end: from planning the assessment programme, to identifying and/or developing appropriate assessment tasks right through to making the assessment judgments.
- It allows for the collection of a number of samples of learner performance over a period of time.
- It can be adapted and modified by the teacher to match the teaching and learning goals of the particular class and learners being assessed.
- It is carried out in ordinary classrooms.
- It is conducted by the learners' own teacher.

- It involves learners more actively in the assessment process, especially if self and/or peer assessment is used in conjunction with teacher assessment.
- It allows the teacher openings to give immediate and constructive feedback to students.
- It stimulates continuous evaluation and adjustment of the teaching and learning programme.
- It complements other forms of assessment, including external examinations and standardised assessments.

As the research questions in the last chapter make clear, the focus of this study is on a comparison of enacted design features against advocated rules within and across SNA and SBA. The literature reviewed in this chapter notes though, that different advocated rules for design are influenced by differing goals relating to the purposes of mathematical assessment. Given that this study is interested in understanding assessment activities at the SBA level in a context of discrepancies between SNA and SBA, the literature reviewed in this chapter is structured to take into account the stated goals within SNA and SBA documentation as a key consideration in understanding the rules that are advocated. The assessment artefacts within SNA and SBA in the empirical sections of this study are considered in relation to the nature of advocated rules within these two assessment systems, which can, in turn, be considered in relation to policy goals.

To get a handle on these matters, mainstream assessment discourses are foregrounded to provide an understanding of motives and goals in SNA that exert influence on the operational elements of SNA artefacts (i.e. the ANA papers in this study). Given the FFL policy mandate for standardisation of assessment, a key argument in this study is that the national policy context provides rules of design that are intended to figure prominently in SNA and SBA, thereby serving as tools for standardisation between the two assessment systems as well as for design of assessment.

In this chapter the literature viewpoints were structured to first give an overview of discourses in assessment. Morgan's (2000) assessment discourse framework, developed to look across national and school-level assessment practices in

mathematics, is particularly useful for understanding assessment in mathematics education. Morgan's work suggests that there are three assessment discourses offering fundamentally different starting points in relation to the design of assessment tasks. Morgan's ideas provide ways of locating design rules in particular discourses on assessing mathematics and to establish possible overlaps of rules between SNA and SBA.

Secondly, literature viewpoints on rules of design for SNA and SBA are also presented. In the SNA literature, these viewpoints are exemplified using examples from a selection of cross national, regional and national programmes. In SBA, viewpoints are considered within research studies associated with classroom assessment. Finally, the literature viewpoints presented allow for a setting up of a literature-based framework against which mathematics assessments within and across SNA and SBA can be critically analysed.

2.2 Discourses in assessment

2.2.1 Overall goals and guiding priorities in assessing mathematics

A key aspect of assessment in school mathematics is about knowing what competencies learners can demonstrate at particular grade level or age. Also by its very nature as a discipline, the goals of mathematics assessment are inextricably intertwined with the broader goals of mathematics education and school education. However, there is undeniably greater emphasis placed on learner achievement in mathematics as an indicator of educational quality than other subjects. In 2000, The Dakar Framework for Action (UNESCO, 2005, p.8) specified the following goal:

> *Improving all aspects of the quality of education and ensuring excellence of all so that recognized and measurable learning outcomes are achieved by all, especially in literacy, numeracy and essential life skills.*

Mathematics as a subject has long enjoyed a privileged status in school curricula worldwide due to its perceived role in science and technology, and subsequent employment prospects. This is reflected by governments' (including South Africa) relatively large investments in improving the quality of mathematics education and extending it to marginalised and underprivileged groups. In South Africa, for

example, the Dinaledi Programme aimed at improving the performance of students in Mathematics and Science in 400 schools; was a focused intervention by the government to improve learner achievement in mathematics (DoE, 2009a).

Jurdak (2009) argued that the first decade of this century witnessed the emergence of a new goal, namely providing equal access to quality education for all students. He explained that this goal was becoming a challenge for the international community, to national governments, schools, educators and parents. Mathematics is considered as a major core school subject and is at the center of this challenge.

The National Council of Teachers of Mathematics (NCTM) assessment standards from the United States offer a summary of goals and standards for mathematics classroom assessment. They argue that standards are about (a) the mathematics, (b) the learning of mathematics, (c) equity and opportunity, (d) openness, (e) inferences, and (f) coherence (de Lange, 1999). The goals of the NCTM standards were to provide a framework to analyse classroom assessment practice and de Lange (1999) argued that they were also aligned to frameworks used in standardised external assessments such as those used by the OECD and IEA in conducting international assessment studies. However, de Lange (1999) was also critical of the goals of agencies and ministries of education underlying mathematical assessments and raised questions about the appropriateness of the mathematics reflected in most traditional tests on the basis that: “the mathematics is generally far removed from the mathematics actually used in real-world problem solving”. Nonetheless, he conceded that there was still much debate over how to define important mathematics and who should be responsible for doing so.

A commonly accepted view is that assessment is always a process of reasoning from evidence. But by its very nature, assessment is imprecise to some degree (Pellegrino, Chudowsky, and Glaser, 2001). Assessment results are only estimates of what a person knows and can do. Every assessment, regardless of whether it features in SNA or SBA, rests on three pillars: a model of how students represent knowledge and develop competence in the subject domain; tasks or situations that allow one to observe students’ performance; and an interpretation method for drawing inferences from the performance evidence thus obtained (Pellegrino,

Chudowsky, and Glaser, 2001).

Other research views on assessing goals in mathematics are related to learner understandings. Studies conducted by Anne Watson from 1998 to 2000 in the United Kingdom had a very clear focus on drawing inferences about learner understandings in mathematics. A key finding for Watson (2000) was that teachers' perceptions about assessing mathematics and the extent to which they applied the assessment were based more on principles and purposes advocated in policy led processes and texts than on their ability to recognise learners' understanding of mathematics.

2.2.2 Morgan's framework on assessment

Morgan (2000) examined the discourses that dominated thinking in England (the country where her research was conducted) and internationally, about assessment in mathematics education by analysing the sets of constructs, assumptions and values that underpin research, curriculum development and teacher education in relation to assessment of mathematics. Morgan (2000) argued that there has been an increasing interest in the role of assessment in the context of curriculum reform amongst researchers as well as curriculum developers. She pointed out that: "assessment methods are not only expected to match the values of the curriculum reform but are also used to coerce teachers into teaching in ways that are consistent with the curriculum objectives". This position is powerful in the South African context of transition to a democratic dispensation, with waves of curriculum reform accompanied by changes in policy goals, curriculum design and the role of teacher.

Morgan's (2000) study identified three mainstream assessment discourses in mathematics that can be applied to understand goals as starting points of mathematics assessment design in environments of curriculum reform. These goals are: psychological; curriculum implementation; and curriculum standards. **Table 2.1** below (from Morgan, 2000, p.67) provides a summary of the main characteristics of each.

Table 2.1: Summary of Morgan's mainstream assessment discourses

	Psychological	Curriculum Implementation	Curriculum Standards
Focus	Individual learner	System-wide curriculum	System-wide outcomes
Aims	To produce valid knowledge about individual learners	To effect reform	To produce higher achievement
Assessment should be	Authentic – identifying real mathematical understanding	Authentic – matching the values of the desired curriculum	Normative and challenging
Learners benefit as a result of	Teaching matched to learning needs	Teaching matched to curriculum aims	Better individual opportunities as the national economy will improve
Teacher's role	To know learners and support their leaning	To implement changes in curriculum and teaching methods	To adopt strategies that will lead to higher outcomes
Learner's role	Learner	Receiver of curriculum	Future worker

The three assessment discourses offer fundamentally different standpoints for teachers to design their assessments. In the psychological discourse the standpoint is the learner and the position advocated by Morgan (2000) is that *design features* involved are directed towards identifying, supporting and knowing about individual learner needs. Morgan (2000) contrasts the individual focus of the psychological discourse to the more policy directed curriculum implementation discourse, where the focus is on all stakeholders implementing curriculum reform according to some standardised curriculum. The role of the teacher is to implement the required changes and transfer knowledge to the learner according to the expectations of the curriculum. This is seen as beneficial to the learner as teaching is closely matched to curriculum aims. The curriculum standards discourse presents a contrast to the previous two discourses, embracing system wide outcomes rather than a narrower focus on individual learners and curriculum implementation. The focus is on efficiency and higher achievement through better economic conditions. Emphasis lifts from the learner and the curriculum to the economy. The teacher is driven to

adopt assessment strategies that will enable the learner to become an efficient future worker.

Morgan's assessment discourse framework provides a reference to better understand assessment motives and goals in mathematics education. Earlier it was indicated that motives and goals have an influence on the design features (or rules) that underlie assessment instruments and the role played by the teacher. This framework provides a route for analysing the influences on rules for assessment in SNA and SBA.

Following from Morgan's aspects, literature on SNA and SBA is considered in relation to advocated rules associated with SNA and SBA and the ways in which design features are conceptualized.

2.3 The design of SNA

2.3.1 Motives

Motives can be seen as broad collective systemic goals derived from strategic plans and policies that lead to certain imperatives and constraints or "advance rules" for design of SNA. In human systems though, with multiple motives somewhat in tension, possibilities for misalignment and contradiction exist. To understand the motives of SNA it is necessary to understand its key purposes. Postlethwaite and Kellaghan (2008, p.3) defined the purpose of national assessments as answers to one or more of the following questions:

a) How well are students learning in the education system (with reference to general expectations, the aims of the curriculum, or preparation for life)?
b) Is there evidence of particular strengths and weaknesses in students' knowledge and skills?
c) Do certain sub-groups of students in the population perform poorly? For example, are there disparities between the achievements of: boys and girls, students in urban and rural locations, students from different language or ethnic groups, and students in different regions of the country?
d) What factors are associated with student achievement? That is, to what extent does student achievement vary with the characteristics of the learning

environment (for example: school resources, teacher preparation and competence, and type of school) or with students' home and community circumstances?

e) Do the achievements of students change over time? And in particular, has student achievement improved, stayed the same, or declined in the time period covered by the introduction of important education reforms?

An ensuing inference on SNA drawn from the purposes listed by Postlethwaite and Kellaghan is that, referring back to Morgan's frame, there are multiple discourses at play that overlap characteristics of psychological, curriculum implementation and curriculum standards strands but there is more emphasis placed on the latter two discourses than on extracting and reporting on psychological information about individual learners. Thus, multiple goals can be identified within stated purposes with the State acting as the key driver of those goals rather than the individual teacher.

Other researchers have suggested that the purpose of national assessments is often not only to establish learner performance within a specific content area (e.g. mathematics) but also, by design, to evaluate the success of policy goals for the education system as a whole. For over 20 years Chile has used assessment as a policy tool, although the mode and purpose of assessment have changed from assessments focused on individual students' outcomes and their futures, to assessments to evaluate schools in a market-driven logic of improvement (Swaffield & Thomas, 2016). Through national assessments, learner achievement levels and the contexts (or reform policies) within which learning takes place 'serve as critical indicators for evaluating the impact of policies in the education system' (Kanjee, 2006). Pellegrino, Chudowsky, and Glaser (2001) suggest that often a single assessment is used for multiple purposes, but note that, in general, the more purposes a single assessment aims to serve, the more each purpose will be compromised and lead to contradictions. For instance, many state tests are used for both individual and program assessment purposes. This is not necessarily a problem, as long as assessment designers and users recognize the compromises, contradictions and trade-offs such use entails (Postlethwaite and Kellaghan, 2008).

2.3.2 Goals

> *National assessments provide a wide range of stakeholders with valuable planning information about the general conditions of schooling and the quality of education.* (Postlethwaite and Kellaghan, 2008)

The literature suggests that there are multiple goals specific to SNA: Five of them drawn from the work of Postlethwaite and Kellaghan (2008) are explained below:

a) Provide data on learner achievement levels

Postlethwaite and Kellaghan (2008) argue that a national assessment is a survey of schools and learners that is designed to provide evidence about the levels of learner achievement in identified curriculum areas (for example, in the areas of reading and mathematics) for a whole education system or for a clearly defined part of an education system (such as Grade 4 students or 11-year-olds). Most researchers who have written on SNA (Postlethwaite & Kelleghan, 2008; Anderson and Morgan, 2008; OECD, 2003; de Lange, 1999) have suggested that the main focus and motive of a national assessment is to describe and evaluate the quality of student learning outcomes that have been produced by schools. But they note that national assessments differ from public (external) examinations – where the main focus is on individual students, certifying their achievement, and selecting them for further education.

b) Provide data on curriculum reform

National assessments also provide important information to stakeholders such as teachers, parents, and the general public on curriculum implementation in the sphere of curriculum reform. Although it has been known for governments to suppress the results of national assessments because of anticipated embarrassment, assessment literature points to the long-term advantages of an open information system (including increased public support for education and a stimulus for reform) as likely to outweigh any short-term disadvantages (Postlethwaite & Kelleghan, 2008). For example the National Centre for Education Statistics[4] (NCES), reported that the State-level National Assessment of Educational Progress (NAEP) administered in the United States of America was an important resource for policymakers and other

[4] The National Center for Education Statistics (NCES), located within the U.S. Department of Education and the Institute of Education Sciences, is the primary federal entity for collecting and analyzing data related to education.

stakeholders responsible for making sense of and acting on state assessment results (de Mello, 2011).

c) Provide data on social and economic objectives, and school-related variables believed to relate to these objectives

A national assessment can have wide-ranging implications for: (a) social and economic policy regarding the overall quality and performance of the education system, including its role in achieving social and economic objectives (for example: equality of opportunity, gender parity, and improving the performance of students from disadvantaged backgrounds); (b) the organization and management of an education system (for example: the provision of public and private education); and (c) learning conditions (for example: instructional time, resources, teacher education, and family support) (Husén, 1987).

d) Provide information for policy decisions

> *Today there is solid consensus among authorities, educators and specialists on the relevance of assessment systems as a guide to educational reforms and, above all, to the adoption of policies to improve the quality of education (Castro, 2000, p.5).*

State institutions often rely on national assessments to inform their decisions on how much students have learned, whether standards are being met, and whether educators are effecting learning (Stiggins, 2002). They are however limited in informing the continuous instructional decisions faced by students and teachers seeking to manage the learning process as it unfolds. Some commentators caution against use of comparative and standardised testing to influence policy decisions. Goldstein and Thomas (2008) argued that there seemed to be very little correspondence between what countries felt are appropriate test items and their actual scores on those items. They reiterate the need for local contextualised testing programmes.

e) Provide an accountability mechanism to rate schools

One of the more explicit but often not declared goals of national assessments is accountability of schools. Information is collected continuously, recorded and

published often in the form of league tables, and performance is also monitored by peer reviews, site visits and inspections. Within all this, there is a high degree of uncertainty and instability and teachers work within a “performative culture” (Ball, 2003). Klecker (2008) argued that the NAEP has since 1969, been the only nationally representative and continuing assessment of what America's students know in various subject areas and together with “The No Child Left Behind Act (2002)” the assessment programme required participation of all schools receiving “Title I” money from federal funds. Similarly in Canada, large-scale SNA are increasingly used in jurisdictions not only to measure student achievement but also to hold schools accountable for the educational outcomes of students (Klinger, DeLuca, and Miller, 2008).

While Morgan’s (2000) psychological discourses feature within the first goal listed above, the other goals associated with SNA are all driven towards her latter two assessment discourses, pointing to the predominance of curriculum implementation and curriculum standards orientations within the SNA literature.

2.3.3 Rules

In this section, an overview of rules (design features) used in SNA cross-national, regional and South African national systemic assessments is provided. Looking across writing detailing the features guiding the development of these assessments allowed me to identify a set of widely used design elements that I was then able to apply to my own analyses of the focal SNA (the ANAs across 2008-2010) and SBA (the three schools’ Grade 6 assessments) in this study.

Literature on national assessment studies (Greaney & Kellaghan, 2008; Postlethwaite & Kellaghan (2008); Ross & Genevois, 2006; and Chinapah, 2003) indicate that there have been various examples of SNA that fall into cross-national, regional or country specific categories. Ross and Genevois (2006) indicated that more than 20 international assessments have been conducted by several agencies during the past fifty years, in a range of subjects and in a large number of both industrialized and developing countries. Well-known SNA studies in mathematics include The Trends in International Mathematics and Science Study (TIMSS), the

Programme for International Student Assessment (PISA), the Monitoring Learning Achievement (MLA) project, Programme for the Analysis of Educational Systems of the CONFEMEN Countries (PASEC), the Latin American Laboratory for Assessment of Educational Quality (Laboratorio) and the Southern Africa Consortium for Measuring Education Quality (SACMEQ).

TIMSS, designed by the International Association for the Evaluation of Educational Achievement (IEA) and PISA by the Organization for Economic Co-operation and Development (OECD) are examples of cross-national studies that have focused on mathematics and science achievement. The MLA, PASEC, Laboratorio and SACMEQ are examples of regional SNA studies. Chinapah (2003) reported that more than 120 countries worldwide have participated in different surveys measuring improvement in learning outcomes in different regions at different grades and in different subject areas. Regional SNA studies are generally confined to countries at more or less the same stage of economic development (Ross & Genevois, 2006).

Some countries (e.g. Brazil Canada, Chile, and the USA) also design their own national assessments with specific local objectives and design features. NAEP (mentioned earlier) for example is a national assessment conducted in the USA. Some of the recent cross national and regional SNA studies in mathematics that South African learners have participated in include the UNESCO designed MLA project, the TIMSS study, and the SACMEQ regional study. National assessments designed by the ministry of education in South Africa include Systemic Evaluations (SE) and the Annual National Assessment (ANA). **Table 2.2** below provides an indication of when these studies were conducted and the targeted grades for participation.

Table 2.2: Examples of SNA in mathematics in South Africa

Study	Year	Target Grade/s
Cross national		
TIMMS*	1999; 2003; 2011	8, 9*
Regional		
MLA	1995; 1999	4
SACMEQ	2000; 2007	6

National		
SE	2000 and 2007; 2004,	3, 6
ANA**	2008; 2009; 2011; 2012; and 2013	1–9

* In 2011, The TIMMS Grade 8 test was taken by Grade 9 learners in South Africa.
** The ANA was written in Grades 3 and 6 in 2008 and 2009; in Grades 1–6 in 2011; and in Grades 1–6 and 9 in 2012 and 2013.

Previous studies have documented that performance by South African learners in these studies is low in comparison to other participating countries that are in the same developmental context (van der Berg & Louw, 2006; Howie, 2006; Kanjee, 2006). Across the studies the mean scores of South African learners are low, usually hovering in the 30% achievement level. A brief description of the key characteristics of the three SNA types relevant to this study is given next.

a) Cross-national studies

International assessments can inform national authorities about the extent to which other school systems "do better" than their own and monitor the evolution over time of their own set of indicators on student outcomes (Grisay & Griffin, in Ross & Genevois, 2006). TIMSS is the most extensive international comparative study to date. In 1995, 41 countries participated and by 2003, the number of participating countries had grown to 50. TIMSS was designed to align broadly with mathematics and science curricula in participating countries. The curriculum model has three aspects: the intended curriculum, the implemented curriculum and the attained curriculum (Reddy, 2006). The results could therefore be used to determine the degree to which learners have acquired the mathematics and science concepts and skills likely to have been taught in school (Reddy *et al*, 2012). The design of TIMSS also allows participating countries, such as South Africa, to compare learner educational achievement across borders and an opportunity to benchmark oneself against other countries.

The assessment framework for TIMSS (Grønmo, Lindquist, Arora, & Mullis, 2013) is organized around two dimensions:

• Content dimension, specifying the subject matter to be assessed; and
• Cognitive dimension, specifying the thinking processes to be assessed.

The table below shows the target percentage of testing time devoted to each content

and cognitive domain for eighth grade assessments.

Table 2.3: Target Percentages of the TIMSS 2011 Mathematics Assessment

Content Domain	Percentage	Cognitive domain	Percentage
Number	30%	Knowing	35%
Algebra	30%	Applying	40%
Geometry	20%	Reasoning	25%
Data and Chance	20%		

The content domains (see Table 2.3 above) are weighted in favour of number and algebra (60%) based on these domains constituting higher proportions of skills to be assessed in many countries' curricula (including South Africa). Each content domain consists of topic areas, and each topic area in turn includes several sub-topics. Across the eighth grade mathematics assessment, each topic receives approximately equal weight in terms of instructional time allocated to assessing the topic. In terms of the cognitive dimensions, the first domain, knowing, covers the facts, concepts, and procedures students need to know, while the second, applying, focuses on the ability of students to apply knowledge and conceptual understanding to solve problems or answer questions. In the TIMSS 2003 framework (Reddy, 2006), knowing was further broken down into knowing facts and procedures (15%) and using concepts (20%). The third domain, reasoning, goes beyond the solution of routine problems to encompass unfamiliar situations, complex contexts, and multi-step problems. Applying is the most targeted (40%) cognitive domain for assessment. Problem solving is central to the applying domain, with an emphasis on more familiar and routine tasks. Grønmo *et al* (2013) argue that TIMSS assesses a range of problem-solving situations within mathematics, with about two-thirds of the items requiring students to use applying and reasoning skills. TIMSS tests incorporated a combination of multiple-choice and open-ended constructed response formats.

b) Regional studies

Regional studies like MLA and SACMEQ are designed specifically for developing countries with similar economic and social development status. Participation in regional studies enables ministries of education to establish how their education

systems are performing at particular time points and to benchmark how their learners are doing in comparison to others in the region. In 1995 The MLA study was a joint UNESCO-UNICEF project where studies of learning achievement were carried out in 40 developing countries in Africa, Asia, the Arab world, the Caribbean, Europe and Latin America. The design of the study was to identify factors that promote or hinder learning in primary schools. The 1999 MLA (Africa) study report (Chinapah et al, 2002) indicated that the results could be used to assess the number of learners at Grade 4 level who had mastered a set of regionally defined basic learning competencies. The MLA mathematics tests were therefore designed to assess the basic knowledge and analytical skills that Grade 4 learners were commonly expected to have.

A major goal of the regional studies like the MLA and SACMEQ is to enhance the assessment capacity of participating countries (Rao & Harshitha, 2004). SACMEQ is a collaborative network of fifteen ministries of education. It was launched in 1995, with assistance from UNESCO's International Institute for Educational Planning (IIEP). Its major focus is on 'capacity building' in the area of educational policy research (Ross & Genevois, 2006). The design of mathematics instruments in SACMEQ is aimed at generating reliable information that can be used by decision-makers to improve the quality of education (Rao & Harshitha, 2004) and identify gaps in learner achievement. SACMEQ research is thus informed by policy concerns identified by Ministers of the SACMEQ member countries.

Van der Berg and Louw (2006) observed from the SACMEQ II study data on mathematics achievement that the educational context in South Africa is quite distinct, even when compared to its regional counterparts. They noted that when achievement in mathematics is considered among learners from schools in different quintiles (i.e. from communities with different poverty indices) (see **Table 2.4** below) South African learners in quintile 5 (more well-resourced) schools scored 155 more points than learners from the quintile 1 (poorly-resourced) schools. Soudien (2008) made the point that SACMEQ II scores revealed that South Africa had the highest difference in mean scores when the extremes (quintile 1 and 5) are compared. In fact there is a distinct difference between all the learners who fall in quintiles 1 to 4 and those in quintile 5 (Soudien, 2008). These findings suggest that it would be

useful to understand and compare the nature of SBA within different socioeconomic school settings.

Table 2.4: SACMEQ II scores for Grade 6 mathematics by quintile and country

Quintile	1	2	3	4	5	Mean
Botswana	491	499	510	508	557	513
Kenya	540	545	555	565	611	563
Lesotho	443	448	448	445	452	447
Malawi	422	427	435	433	447	433
Mauritius	519	564	587	620	640	584
Mozambique	526	525	531	530	538	530
Namibia	403	402	411	425	513	431
Seychelles	520	541	555	576	579	544
South Africa	**442**	**445**	**454**	**491**	**597**	**486**
Swaziland	506	511	511	513	541	517
Tanzania	484	511	529	528	560	522
Uganda	484	497	498	509	543	506
Zambia	414	425	436	434	466	435
Zanzibar	478	472	478	479	484	478
Mean	**468**	**480**	**485**	**492**	**560**	**468**

(Source: van der Berg and Louw, 2006)

The SACMEQ mathematics test frameworks cover curriculum 'topics' that are common across the member systems. Drawing from an audit of country specific curriculum topics, the design of test instruments is based on content domains and a hierarchy of skills. In the SACMEQ II Project "mathematics literacy" was defined as "the capacity to understand and apply mathematical procedures and make related judgments as an individual and as a member of the wider society" (Ross et al, 2005). The SACMEQ designers used the IEA domains as a beginning point for an extensive investigation of curricula, textbooks, and examinations for Grade 6 pupils within SACMEQ school systems and came up with the following three domains:

- Number: Operations and number line, square roots, rounding and place value, significant figures, fractions, percentages, and ratios.
- Measurement: Measurements related to distance, length, area, capacity, money, and time.

- Space-Data: Geometric shapes, charts (bar, pie, and line), and tables of data.

Against these three domains, a total of five hierarchical mathematics skill levels were identified for the mathematics test (Ross *et al*, 2005):

- Level 1: Pupils at this level should be able to identify simple shapes and link simple patterns and shapes to simple digits, to recognize units of measurement, to name basic shapes, and to undertake simple single operations using up to two-digit numbers.
- Level 2: Pupils at this level should be able to recognize simple fractions in both numerical and graphical forms, to identify data presented in tables, to make basic calculations using simple measurement units, and to understand numeration with simple computations.
- Level 3: Pupils at this level should be able to extend and complete number patterns, to translate shapes and patterns, and to convert measurement units when making simple single-step calculations.
- Level 4: Pupils at this level should be able to combine operations in order to link information from tables and charts in performing calculations, to apply two or three-step number operations applied to measurement and conversion problems, and to identify and use appropriate information in the subsequent steps of a calculation.
- Level 5: Pupils at this level should be able to make calculations and interpretations linking data from tables and graphs, and to make computations involving several steps and a mixture of operations using fractions, decimals, and whole numbers.

The final construction of the mathematics test was done by combining the mathematics skill levels with mathematics domains to develop a test blueprint where test items were matched to a particular domain and skill level. All items were presented in multiple-choice format.

c) National assessment studies

The response of the South African government to low achievement levels in cross national and regional SNA has been to conduct regular national assessments at key

stages. Grisay and Griffin in their analysis of cross-national studies claim national assessments are better able than international assessments to provide information on specific characteristics of a school system on: a) whether all aspects of the curriculum have been covered, and b) the proportion of students that meet specific national standards (Ross & Genevois, 2006). In South Africa, since the democratic transition in 1994, the national education ministry has made use of two types of national assessments to report the results of learner performance. The first type involved systemic evaluations (SE), conducted on a 5 year cycle on a random sample of public schools at either grade 3, 6 or 9. The results were used to report on the policy goals of access, equity and quality as indicators of the 'health' of the education system (DoE, 2005a). Hence the focus of the SE studies focused on a curriculum implementation and curriculum standards discourse and reported on contextual factors influencing learner achievement without substantive attention towards the psychological elements of learners' performance in mathematics The 2004 grade 6 SE study conducted by the DoE highlighted the central role of the educator and principal in raising the quality of education (Kanjee, 2006) but these evaluations did not look specifically at the nature of assessment practices as a key part of the role of the educator in raising quality.

Following on from the designs of the cross national and regional SNA, the SE studies in South Africa were devised on the basis of assessing learner performance in mathematics and languages on competencies stipulated in the national curriculum statement and revisions. Many features of the TIMSS and SACMEQ methodology on content domains, item type and skills levels can be seen in the design of the SE studies. The SE assessment covered the breadth of the then South African curriculum utilising questions based on a framework that included five content domains (number, patterns, shape and space, measurement and data). In South Africa, these content domains were referred to as learning outcomes at the time. The use of cognitive levels such as knowledge, application and problem solving were similar to those used by TIMSS. As in TIMSS, the item types were a combination of multiple-choice and open ended questions.

An interesting feature of the SE framework was that difficulty level was considered separate from cognitive level, using a three level category scale of easy, moderate

and difficult differentiated on the basis of the grade level of the curriculum. For example a Grade 6 mathematics question would be considered easy if it was based on skills extracted from the Grade 5 mathematics curriculum. The specimen test framework used in the Grade 6 systemic evaluation study conducted by the DBE in 2004 is indicated in **Figure 2.1 below**:

LO	AS	N. of items	Item type		Difficulty levels			Cognitive levels to "tap into"		
			MCQ	OEQ	Easy	Moderate	Difficult	Know.	Application	Problem solving
		…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…
...	…	…	…	…	…	…	…	…	…	…
Total items		X=all items	X_1= all MCQs	X_2= all OEQs	…	…	…	…	…	…
Key: LO = Learning outcome; AS = Assessment standard; MCQ = Multiple choice question; OEQ = Open-ended question (or free-response question); Know. = Knowledge										

Figure 2.1: Specimen of a "test framework" from the Grade 6 SE study

The second type of national assessment, referred to as the ANA, was underlain by a similar framework and targeted a more diagnostic (psychological) interpretation of learner achievement (DBE, 2014). Elements of the SE framework established from policy prescripts and DBE reports had a significant bearing on the design of the ANA tests. For example, the 2008 ANA tests were based on a framework with similar features to those included in the SE studies.

The ANA, conducted for the first time in 2008, was aligned to an educational intervention campaign, the FFL. It was not compulsory for all schools to participate. By 2011, the state had included ANA as an indicator to measure quality learning outcomes in its education sector plan (DBE, 2010) and has since evolved into a "a census type of external assessment involving all learners in all public schools" (DBE, 2014, p.8). A similar kind of census testing is done in Chile among fourth and eighth grade learners (Kanjee & Braun, 2006). In 1998, Chile introduced a set of national assessments known as the "Sistema de Medición de al Calidad de la Educación" or

SIMCE which annually tests hundreds of thousands of children on their knowledge in language, mathematics, social and natural sciences (Swaffield & Thomas, 2016). In 2011, the ANA targeted a population of around 5 million learners, and has since grown into testing almost 7 million learners (DBE, 2014), a rare feat in international assessment. In South Africa, the ANA is one of the "largest initiatives undertaken to improve learning and teaching in recent years" (Kanjee & Moloi, 2014, p.108), and like Chile has been one of few countries to develop and implement national assessment and information systems in education on this scale.

2.3.4 Summary points

From the above discussion on SNA, the following summary points are noted:

1). The goals associated with a particular mathematics assessment discourse have implications for how one looks at other assessment foci such as the coverage of content, assessment processes and hierarchies, and item format that feature in the makeup of assessment designs in SNA through particular viewpoints.

2) There are multiple goals attached to SNA and these include providing: a) data on learner achievement levels, b) data on curriculum reform, c) data on social and economic objectives, d) information for policy decisions, and e) an accountability mechanism to rate schools. Each goal fits with different assessment discourses in Morgan's (2000) framework.

3) In South Africa, participation in SNA studies (e.g. SACMEQ) has shown that the school setting can be an influential factor in the achievement of learners. Therefore assessment performance in different types of school settings (such as township, inner-city and suburban) and in different quintiles assessment widely differs.

4) Common design features in cross national, regional and national assessments include focus on content domains for mathematics, item type and a hierarchical structure of cognitive skills. In the SE studies (and later the ANA) conducted in South Africa, difficulty levels are distinguished from cognitive demand.

2.4 The design of SBA

2.4.1 Motives

Literature on school-based assessment (SBA) suggests it is an assessment embedded in the teaching and learning process. Hence its focus is generally on the

delivered curriculum in contrast to the make-up of SNA which focus more on intended skills and coverage. However, the purpose of SBA must be, and must be seen to be, of major relevance to the outcome of the high stakes assessment such as those associated with SNA (Long, 2006). Braun and Kanjee (2006) argue that the most common type of assessment is school-based. As mentioned already, these assessments are usually devised and administered by class teachers, although some are the work of the school principal or other teaching staff, and they carry a number of important characteristics (Davison & Hamp-Lyons, 2009) which distinguish them from SNA.

In many educational systems, such as those of Australia, Canada, the United Kingdom and Finland, SBA is used extensively or exclusively to provide information about student achievement. In Hong Kong, SBA has been a part of the public examinations system and since 1978 has gradually become a core component of the Hong Kong Certificate of Education Examination (HKCEE) (Davison and Hamp-Lyons, 2012). Over the last 40 years, both Finland and Sweden have shifted from highly centralized systems emphasising external testing to more localised systems using multiple forms of assessments (Darling-Hammond, 2008). Great Britain makes use of a combination of external and school-based tasks based on the national curriculum and course syllabi.

In South Africa, the policy context for SBA in the primary schools is regulated through the National Education Policy Act of 1996*, Assessment policy in the General Education and Training Band (GET): Grade R to 9 and ABET* (DoE, 1998), the Revised National Curriculum Statement (RNCS)[5] (DoE, 2002b) and the National Protocol on Assessment for schools in the General and Further Education bands: Grade R–12 (DoE, 2005d). SBA includes regular classroom tests and end of year examinations. SBA tasks are devised and administered by teachers and have the dual purpose of improving learning in the classroom and promoting learners to the next grade (Buhlungu *et al*, 2007).

2.4.2 Goals

[5] In this study the national curriculum is referred to as the Revised National Curriculum Statement (RNCS) introduced in 2002. In 2013, the RNCS was replaced with Curriculum Assessment Policy Statement (CAPS).

Researchers present many different goals for SBA and as in the case of SNA they can be linked to Morgan's (2000) assessment discourse framework. Darling-Hammond (2008) and Long (2006) suggest that a key goal for SBA is integration: "The integration of curriculum, assessment, and instruction in a well-developed teaching and learning system creates the foundation for much more equitable and productive outcomes" (Darling-Hammond, 2008, p.8). This view suggests that there is often an intention in SBA to have goals that straddle all three assessment discourses, namely, psychological, curriculum implementation and curriculum standards, although the enacted discourse might show up goals that belong mainly to a particular discourse with less emphasis on the other two. Long (2006) argues that it is widely recognized that traditional tests and examinations in SBA have goals mainly directed towards curriculum implementation and need to be complemented by alternative ways of assessing (psychological) what learners know and can do. SBA offers an opportunity to introduce these alternative assessments.

Other researchers suggest that it is accepted that SBA, through classroom tests and assessments, play a central role in the evaluation of student learning and therefore focus on a more psychological discourse. They make expected learning outcomes explicit to learners and parents and show what types of performances are valued. The validity of information they provide, however, depends on the care that goes into planning and preparation of tests and assessments (Linn & Miller, 2004). SBA, in this view, should be an integral part of the teaching and learning process, and there should be a mutual influence (de Lange, 1999).

In further support of a psychological emphasis, Linn & Miller (2004) argue that the main goal of classroom testing and assessment is to obtain valid, reliable, and useful information concerning learner achievement, while minimising the influence of irrelevant or ancillary skills. They argue that the following basic steps in classroom testing and assessment would significantly contribute towards the goal of improved learning and instruction:

a) Determine the purpose of assessment,

b) Develop specifications,

c) Select appropriate assessment tasks,

d) Prepare relevant assessment tasks,

e) Assemble the assessment,

f) Appraise the assessment, and

g) Use the results.

In order for this cycle to be followed, teachers need to be aware of the connections between the tests, tools and curricular goals and how to generate relevant feedback from the test results (de Lange, 1999). Some identify classroom assessment feedback with formative assessment (Biggs, 1998) but de Lange (1999) and Black and Wiliam (1998) agree that formative and summative assessments are not mutually exclusive. Summative assessment in the form of end-of-year tests can give teachers evidence of how well they handled the formative assessments given during the course of the year, assuming that the underlying philosophy is coherent and consequent. This leads some researchers to argue that the differences between formative and summative assessment within the classroom are more related to timing and the extent of accumulation than anything else (de Lange, 1999). In the psychological discourse, it is viewed as important that assessment is criterion-referenced, incorporating the curriculum and resulting in aligned assessment practice (Linn. And Miller, 2005)

Withers (2005) points out that multiple assessment discourses may be at play since the real objectives for any classroom test come directly from considering two things in conjunction: the first is the actual educational context (e.g. curriculum reform or curriculum standards) in which the results of the testing will be used; the second is the knowledge and understanding (psychological) the learner is expected (or able) to bring into the test room. The objectives determined in the SBA are local and specific, and determining where learners are is the starting point for determining what the test and its items should look like (ibid, 2005). Establishing the starting point of the assessment design of teachers involved in this study was significant in understanding the make-up of the assessment tools used in their SBA.

2.4.3 Rules

In many cases, local school based assessments complement centralised "on-demand" tests, constituting up to 50% of the final examination score. Tasks are mapped to the standards or syllabus for the subject and are selected because they represent critical skills, topics, and concepts. They are often outlined in the curriculum guideline, but they are generally designed, administered, and scored locally, and may be based on common specifications and evaluation criteria. Whether locally or centrally developed, decisions about when to undertake these tasks are made at the classroom level, when appropriate for students' learning process. Teachers can get information and provide feedback as needed; something that traditional standardised tests cannot do (Darling-Hammond, 2008).

The earlier discussion on SNA linked design features to a framework or blueprint that designers (either locally or from an international agency) used. These features included content domain, difficulty level, cognitive demand and item type. The literature on SBA indicates that these reference points continue to be relevant but may be configured differently in SBA. In the discussion that follows I pick up on three SBA design approaches and features that were relevant for the context of this study, noting the ways in which they differ from SNA specifications in the SBA context.

a) The table of specifications

To ensure that SBA measures a representative sample of instructionally relevant tasks, it is important to develop specifications that can guide the selection of test items and assessment tasks (Linn & Miller, 2004). These authors suggest that one useful and widely used device for this purpose is a two-way chart, called a table of specifications (or test blue print). Such tables or blue prints involve: a) preparing a list of instructional objectives, b) outlining the course content, and c) preparing the two-way chart. A table of specifications can be used in SBA and SNA. SNA test designers like those working in UNESCO's International Institute for Educational Planning (IIEP) refer to their two-way chart specification as a test matrix (Withers, 2005). In SNA, the matrix is completed by a panel of experts consisting of policy experts, curriculum experts and those who will eventually develop the test. In SBA, this view suggests that the matrix is completed by the teacher as the test designer with the help of a senior teacher or moderator.

The list of objectives in a two-way chart or matrix is limited to those outcomes that can be measured by classroom assessment and describes the types of performance that students are expected to demonstrate. The instructional content refers to the area (or skill) in which each type of performance is to be shown. The two-way chart relates the instructional objectives to the instructional content, thus specifying the nature of the sample of test items and assessment tasks.

An example of a two-way chart is indicated in table 2.5 below. The table is constructed by:

1. Listing the general instructional objectives across the top of the table
2. Listing the major content areas down the left side of the table.
3. Identifying the proportion of the test items that are devoted to testing each objective and each content area.

Table 2.5: A mathematics two-way chart on Fractions relating instructional objectives to instructional content

Content Area	Objectives			Total items
	Adds proper fractions	Adds proper fractions and mixed numbers	Adds mixed numbers	
Denominators are alike	5	5	5	15
Denominators are unlike (with common factor)	5	5	5	15
Denominators are unlike (without common factor)	6	7	7	20
Total number of items	16	17	17	50

Source: Linn and Miller (2004): Table of specifications for a 50-item test on addition of fractions

The final distribution of items in the table of specifications ideally reflects the emphasis given during instruction. Objectives considered more important by the teacher are allotted more test items. Linn and Miller (2004) point out that although the decisions involved in making the table are somewhat arbitrary, and the process often time consuming, the preparation of a table of specifications is one of the best means to ensure that the total set of test items measures a representative sample of instructionally relevant objectives.

The table of specifications applies not only to items within a classroom test but also to more medium term performance assessment tasks. The weight and distribution given to the performance of such an assessment task should reflect the importance of the objective. A table of specifications for a performance assessment task (e.g. summative assessment at the end of a term) would need to be broader than indicated in the table above, especially if the requirement is to test a broader range of instructional objectives related to a content area or section of work. Table 2.6 is an example of a table of specifications designed to test cognitive skills and content areas on fractions and decimals.

Table 2.6: A mathematics table of specifications on Fractions and Decimals

Content Area	Objectives			Total items
	Procedural skills	Understanding	Application	
Simple fractions	5	10	5	20
Mixed numbers	5	15	10	30
Decimals	5	10	5	20
Decimal-fraction relationships	10	10	10	30
Total number of items	25	45	30	100

Source: Linn and Miller (2004): Table of specifications for fractions and decimals

At this 'topic' or 'unit of work' level, we start to see test objectives that link with the cognitive demand aspect that featured in SNA. But some differences in the test configuration between SNA and SBA start to emerge. The content area is limited to one or two specific topics. While this is not as broad as the content coverage for SNA studies which cover three to four different content domains, within a specific content area the assessed skills cover a range of cognitive demand levels. A teacher using this kind of specification table can design a classroom assessment (e.g. on fractions and decimals) with several more items tapping into a range of cognitive demands (or test objectives) than would be possible with the traditional time bound standardised tests associated with SNA.

b) The Taxonomy for Learning, Teaching and Assessing (LTA)

Literature on tapping into a broad range of instructional test objectives in classroom

learning contexts stems from the work done by Benjamin Bloom (1956) in his seminal work on the Taxonomy of Educational Objectives. An in-depth analysis of test objectives is also given in Anderson's (2005) study of objectives, evaluation and the improvement of education. The purpose of Anderson's study was to revise Bloom's (1956) Taxonomy of Educational Objectives. Bloom's Taxonomy divided educational experiences into two domains: the cognitive and the affective. Within the cognitive domain, Bloom pointed to six major categories: knowledge, comprehension, application, analysis, synthesis and evaluation (Bloom et al, 1956). Anderson's revised Taxonomy was published as the Taxonomy for Learning, Teaching and Assessing (LTA) (Anderson *et al*, 2001). Key features of the revised Taxonomy are indicated in the figure below. Withers (2005) argued that such taxonomies provide elements to structure test specifications.

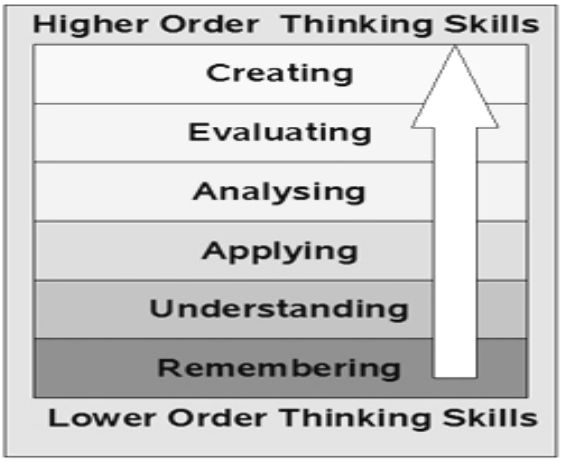


Figure 2.2: Anderson's revised Taxonomy of Educational Objectives

Highlighting differences between the original and revised taxonomies is not a central focus of this study. A substantive analysis of the taxonomies can be found in (Anderson *et al*, 2001). Of greater significance to this study were the implications that Anderson's study had for assessment design. Anderson (2005) wanted to establish, through a revised taxonomy, a strong connection among objectives, assessment and instruction.

The LTA Taxonomy proposed a two-dimensional assessment framework to understand educational objectives. Along the horizontal axis was the cognitive process dimension which ranged across: 1) remember, 2) understand, 3) apply, 4) evaluate, and 5) create. The vertical axis was the knowledge dimension, considered as the lowest level of Bloom's Taxonomy, crossed all levels of the cognitive

dimension. It consisted of four general types of (hierarchical) knowledge: factual, conceptual, procedural and metacognitive (Anderson, 2005).

Factual knowledge is considered knowledge that a student must know about a subject matter. Conceptual knowledge is knowledge of classifications and categories, principles and generalisations, and theories, models and structures. Procedural knowledge is considered as knowing how to do something. It includes methods, techniques, algorithms and skills. Metacognitive knowledge is knowledge of cognition in general as well as awareness of one's own cognition (Anderson, 2005). The LTA taxonomy framework is indicated in table 2.7 below.

Table 2.7: The LTA Taxonomy

	The cognitive process dimension					
The knowledge dimension	1. Remember	2. Understand	3. Apply	4. Analyse	5. Evaluate	6. Create
a) Factual knowledge						
b) Conceptual knowledge						
c) Procedural knowledge						
d) Meta-cognitive knowledge						

Source: Anderson (2005): The Taxonomy Table

Anderson (2005) argued that the revised taxonomy had the following advantages:

1) Prototypes of assessment tasks increased the validity of the assessment tasks, and
2) It increased the efficiency of the preparation of the assessment tasks

Anderson (2005) argued that all assessment tasks are derived from the same blueprint but different objectives require different approaches to assessment. The LTA taxonomy encourages teachers to develop prototypical assessment tasks for the various cells of the table.

c) The assessment pyramid

de Lange (1999) argued that the idea of teachers using a framework to design assessments was not commonplace and a neglected aspect of classroom

assessment. He noted that efforts to design and describe frameworks in mathematics assessment, most notably, the TIMSS monograph, Curriculum Frameworks for Mathematics and Science (Robitaille et al., 1993) and the Measuring Student Knowledge and Skills: A New Framework for Assessment (Organization for Economic Cooperation and Development [OECD], 1999) focused mainly on SNA because of the emphasis in some countries (like the USA) on using standardised tests.

de Lange (1999) argued that a framework for designing classroom assessment must give a structured indication of what the key concepts to consider in the assessment of mathematics are and must bring into focus an understanding of the features that figure in the design of mathematics assessment activities. In the context of large-scale SNA type assessment, the methodology used is usually a statistical model that characterizes expected data patterns, given varying levels of student competence. But in SBA, the assessment methodology is usually intuitive or qualitative rather than based on formal statistical models.

He suggested the following principles be taken into consideration by teachers when planning and designing their assessments:

1) The main purpose of classroom assessment is to improve learning
2) The mathematics is embedded in worthwhile (engaging, educative, authentic) problems that are part of the students' real world.
3) Methods of assessment should be such that they enable students to reveal what they know, rather than what they do not know.
4) A balanced assessment plan should include multiple and varied opportunities (formats) for students to display and document their achievements.
5) Tasks should operationalize all the goals (cognitive demand) of the curricula (not just the "lower" ones). Helpful tools to achieve this are performance standards, including indications of the different levels of mathematical thinking.
6) Grading criteria should be public and consistently applied; and should include examples of earlier grading showing exemplary work and work that is less than exemplary. In this way difficulty can be linked to the curriculum grade.
7) The assessment process, including scoring and grading, should be open to

students.

8) Students should have opportunities to receive genuine feedback on their work.

9) The quality of a task is not defined by its accessibility to objective scoring, reliability, or validity in the traditional sense but by its authenticity, fairness, and the extent to which it meets the above principles.

de Lange (1999) presented an *assessment pyramid* (see **Figure 2.3** below) with three interconnected dimensions to describe the assessment process. The three dimensions are a) the content domains of mathematics, b) the three levels of mathematics thinking and understanding and c) the difficulty of the questions posed. The dimensions are not meant to be orthogonal, and the pyramid is meant to give a fair visual image of the relative number of items required to represent a student's understanding of mathematics.

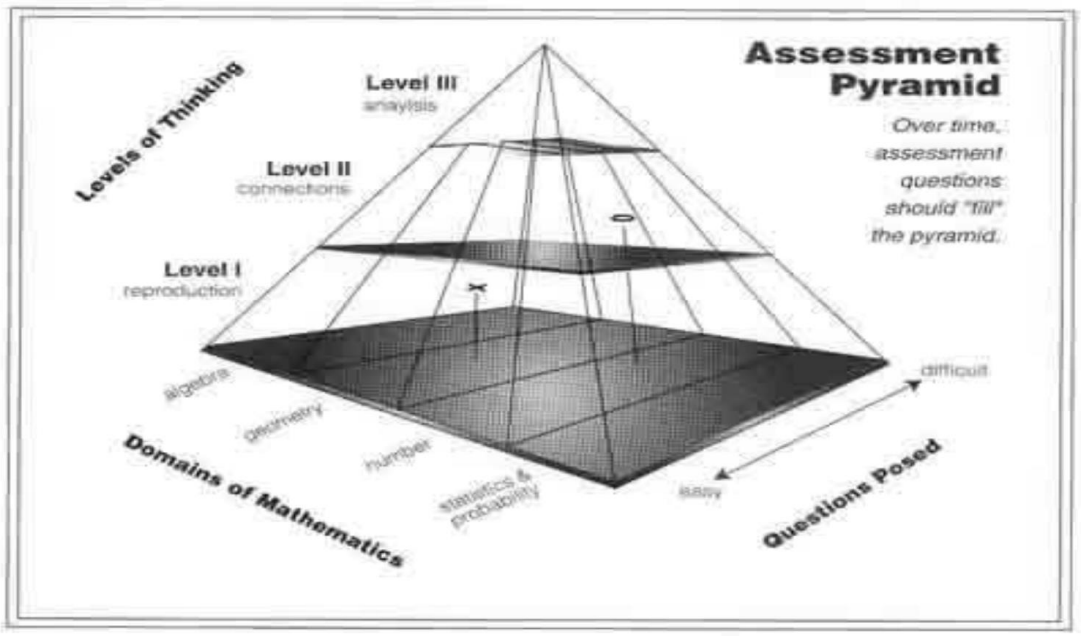


Figure 2.3: de Lange's assessment pyramid

The content domains of mathematics were listed as: Algebra, Geometry, Number and Statistics and Probability. The three levels of thinking were listed as: reproduction, connections and analysis. de Lange (1999) argued that because we need only simple items for the lower levels, we can use more of them in a short amount of time. For the higher levels we need only a few items because it will take some time for the students to solve the problems at this level. The difficulty level

ranged from easy to difficult and was linked to question complexity. de Lange (1999) argued that the assessment pyramid was a model for classroom assessment that embraced many elements of external assessment (SNA) frameworks such as those used by the OECD designed for the PISA studies. An important philosophy of his research was to connect internal and external assessment frameworks as much as possible.

2.4.4 Summary points

1) SBA has a number of important characteristics that makes it different to SNA. Generally, SBA is aligned with the delivered curriculum and covers a greater range of topics than is usually the case with SNA.
2) Some countries have shifted away from centralised assessments and have focused more on supporting an efficient and well-structured SBA in schools while others, through the introduction of more standardised testing, have neglected teacher training on assessment.
3) Like SNA the goals of SBA can be linked to multiple discourses, but emphasis in the literature is on the psychological discourse.
4) Well-structured SBA has specifications that guide the selection of test items and assessment tasks with clear instructional objectives. For example, a two-way chart relates the instructional objectives to the instructional content, thus specifying the nature of the sample of test items and assessment tasks.

Overall, it is important to point out that there are overlaps in the literature between SNA and SBA in terms of design aspects. Notably, there are aspects of SNA like mathematical domain and cognitive demand that feature in SBA. Difficulty level is also considered in SNA studies (e.g. SACMEQ and Systemic Evaluations) and there are SBA studies where difficulty level is judged empirically by teachers as well. A positive interaction between SNA and SBA is viewed as a desirable goal in many country contexts. It is the premise of this study to unpack and explore this interaction in the South African context.

The exploration in this study focuses on the design features underlying SNA and espoused for SBA by teachers in the three school settings, and the enacted design features seen from analysing the nature and range of tasks (test items) selected

within SNA and SBA contexts.

2.5 Categories identified in SNA and SBA for analysing test items

From the literature discussion on SNA and SBA categories for assessing mathematics can be wide and varied and can include a broad range of ideas from curriculum overview, content areas, level of competencies, classification levels, taxonomies, cognitive strands, complexity of thinking skills, depth of knowledge, difficulty range, everyday contexts, formats, feedback, grading, and coherence and balance (de Lange, 1999; Webb, 2002; Anderson, 2005) and many more. In this study, the following operational elements are identified from the literature as having a specific focus or influence on the design features that are included in Mathematics assessments in SNA and SBA contexts. These aspects are summarised as mathematical coverage, difficulty level, cognitive demand, language demand, and item format. Each one of them is explained further below.

2.5.1 Mathematical coverage

Mathematical coverage is viewed as an essential assessment feature of both SNA and SBA. Since it entails what is to be assessed, it is a fundamental feature and usually the first consideration in assessment design. The literature on design features of SNA and SBA suggests that there may be different configurations on how mathematical coverage is applied in SNA and SBA contexts. Considerations in this regard are the selections of one or more broad mathematical domain or content area (e.g. number) or organising field and the more in-depth selections that apply to the assessment of a specific topic or skill within a content area (e.g. fractions)

Within the broader mathematics education community, there is general agreement on what constitutes a mathematical field or content domain. de Lange (1999) made the point that the mathematics community has for some time chosen to organise the content of the relevant mathematics around a few “big ideas” or “themes.” His research on the ‘assessment pyramid’ listed the generally accepted content domains as Algebra, Geometry, Number, and Statistics and Probability.

In the SNA studies mentioned in this chapter, similar content domains were

emphasised as organising fields of mathematical content. The content dimension in the TIMSS assessment framework (Grønmo, Lindquist, Arora, & Mullis, 2013) listed the selected strands as: Number, Algebra, Geometry, Data and Chance. The SACMEQ designers (Ross *et al*, 2002) listed 3 domains namely: Number, Measurement: and Space-Data in their test blue print.

Generally, the literature on SBA suggests that Mathematics school curricula are usually organized into similar strands to those used in SNA but the number of strands selected may be different. de Lange (1999) argued that mathematics school curricula are organized into strands that classify mathematics as strictly compartmentalized disciplines and often with an over-emphasis on computation and formulas.

In curriculum documentation in South Africa, mathematical coverage is a concept used to establish what mathematical skills or content learners are taught at a particular grade level. Within South Africa's outcomes based framework (OBE), drawn from the mandated curriculum that was in place at the time of the empirical data collection in this study, the curriculum content was structured according to learning outcomes (LOs) and assessment standards (ASs)[6]. From the literature on SNA content domains are similar in meaning to learning outcomes.

A learning outcome (LO) was a description of what (knowledge, skills and values) learners should know, demonstrate and be able to do at the end of the General Education and Training (GET) band. A set of LOs aimed to ensure integration and progression in the development of concepts, skills and values through the assessment standards. Five mathematical domains were specified that teachers and learners had to engage within the general education and training (GET) band. These were similar to those indicated by de Lange (1999) and were listed in the RNCS (DoE, 2002b, p.6) as:

a) <u>LO 1: Numbers, operations and relationships</u>

The learner will be able to recognise, describe and represent numbers and their

[6] In the 2013 curriculum and assessment policy statement (CAPS), the terminology on learning outcomes and assessment standards are referred to as content areas and the general content focus.

relationships, and to count, estimate, calculate and check with competence and confidence in solving problems.

b) LO 2: Patterns, functions and algebra

The learner will be able to recognise, describe and represent patterns and relationships, as well as to solve problems using algebraic language and skills.

c) LO 3: Shape and space (geometry)

The learner will be able to describe and represent characteristics and relationships between two-dimensional shapes and three-dimensional objects in a variety of orientations and positions.

d) LO 4: Measurement

The learner will be able to use appropriate measuring units, instruments and formulae in a variety of contexts.

e) LO 5: Data handling

The learner will be able to collect, summarise, display and critically analyse data in order to draw conclusions and make predictions, and to interpret and determine chance variation.

Each LO was comprised of assessment standards. Assessment standards aimed to describe the level at which learners should demonstrate their achievement of LOs and the ways (depth and breadth) of demonstrating this achievement. They were grade specific and showed how conceptual progression was delineated in the curriculum. They embodied the knowledge, skills and values required to achieve the LOs (DoE, 2002b). Each AS was therefore a statement of expected competence. Two examples of an AS are listed below.

Example 1

Grade 6 LO 1 - AS 1: *counts forwards and backwards in decimals*.

Example 2:

Grade 6 LO 4 - AS 9: *investigates and appropriates (alone or as a member of a group):*

a) Perimeter using rulers or measuring tapes.

b) Area of polygons (using square grids and tiling) in order to develop rules for

calculating the areas of squares and rectangles.

c) *Volume/capacity of objects in order to develop rules for calculating the volume of rectangular prisms.*

The structure of an AS varied. In Example 1, the sub-skills of counting decimals are combined in a single statement. In Example 2, the AS is structured to have 3 distinct sub-skills dealing with perimeter, area and volume respectively. In such cases, the learner was expected to be assessed on each sub-skill in SBA but the method for doing so was not prescribed in the RNCS.

Altogether 47 ASs were listed in the mathematics curriculum across 5 learning outcomes (DoE, 2002b) in Grade 6. **Table 2.8** below gives a breakdown of the number of ASs and sub-skills that featured in each learning outcome.

Table 2.8: Number of ASs per LO in the RNCS

LO	Number of ASs	Number of sub-skills of ASs
LO 1: Number	12	38
LO 2: Patterns	6	15
LO 3: Shape and space	8	13
LO 4: Measurement	11	19
LO 5: Data	10	13
Total	47	98

The number of sub-skills attached to an AS gives a more complete picture of the anticipated work and competencies to be assessed within each LO. In the above table, it is clear that LO1 that deals with numbers, operations and relationships has the most skills for learners to be assessed on within the overall basket of mathematical competencies. The breakdown of the RNCS into LOs and ASs suggests that the assessment tasks of South African teachers would be weighted more strongly in terms of LO 1 (Number) than the other content domains. This is in keeping with content domain proportions indicated earlier for cross national studies such as TIMSS which also weighted its content domains in favour of number and algebra (60%) based on these domains constituting higher proportions of skills to be assessed in many countries curricula (including South Africa).

However it should be flagged that in SBA contexts, the weighting of mathematical content domains might be reflected differently as school based tasks often focus on specific topics and the configuration might resemble more the structure of the two-way tables of Linn and Miller suggested earlier. At the single test level, SBA artefacts may include fewer ASs but with more individual items than commonly represented in SNA. The focus of this study was to unpack in detail the structure of tasks and overall assessments among selected teachers and compare how mathematical coverage was configured in SNA and SBA contexts. In the analysis of SNA and SBA, frequency item level counts similar were applied to LOs and ASs and compared to the figures listed in **Table 2.8**.

2.5.2 Difficulty level

In the SNA discussion on national assessments in South Africa, I noted that in the design frameworks of the SE studies and the ANA, difficulty level is listed as an assessment feature. de Lange (1999) also flagged this as a feature of the assessment pyramid. By considering the level of difficulty of individual questions posed in assessment tasks, the overall spread of difficulty as well and the levels that teachers select for their assessment tasks can be established. de Lange's principles of designing assessment (and evidence outlined in Chapter 1) suggest that teachers often design assessment tasks that target minimal levels of achievement by learners in a specific grade. Therefore interpretations of difficulty level are important to make sense of such value judgements. The literature offers a number of different ways to consider the notion of difficulty.

An early view on difficulty level can be found in de Lange's (1999) work on establishing an assessment pyramid for constructing mathematics tests. The *assessment pyramid* mentions difficulty level as ranging hierarchically from easy to difficult and links are made to hierarchical levels of thinking (cognitive demand) and content domains of questions posed in assessment tasks. de Lange (1999) implied that more cognitively demanding a question, the more difficult it was and suggested that an assessment tool should contain items of varying difficulty levels ranging from easy to difficult.

More recently, State departments (like the DBE) and assessment agencies involved in SNA (e.g. IIEP and the IEA) are increasingly looking towards more statistical models for understanding difficulty level. One method of test analysis used widely in SNA is item-response theory (IRT), One specific IRT model used in large scale assessment studies is Rasch analysis. By applying a Rasch analysis, test designers link the range of skills and competencies that learners display in responding to the test questions to difficulty level. In simple terms, a Rasch analysis provides a "measured" link between "scores" and "skills" (Ross, 2005) to determine the difficulty levels of test items in ways that are based on empirical datasets of performance. Unlike descriptive raw scores (average percentages) that describe performance in the entire test, Rasch scores measure performance on each item that a test candidate attempts successfully. Consequently, they reveal "attributes" of candidates which would otherwise not be known through average or "classical test theory" scores (Ross, 2005).

Proponents of IRT models suggest that the difficulty of the test depends on its purpose (Anderson and Morgan, 2008). If the purpose is to monitor the performance of all students in the target population, then the distribution of difficulty of the test items should match the distribution of achievement of the target population. As a general rule, IRT analysts suggest that two-thirds of a test should consist of items that two-thirds of the population have between a 30 and 70 percent likelihood of answering correctly. The other third of the test should be evenly divided between items that more than 70 percent of students taking the test are likely to answer correctly and items that fewer than 30 percent are likely to answer correctly (Anderson and Morgan, 2008). However, due to the technical nature of applying IRT methods, it is seldom used by school teachers in SBA, where the literature points to evidence of more empirical ways of defining difficulty levels.

Interpretations on difficulty that are more closely related to teacher development/use of SBA can be found in the work of Leong (2006). He claimed that while there were a host of construct validation procedures (such as IRT) to aid item writers in ensuring that test items measure the construct they are intended to measure, there is a general lack of information in the literature on how to vary the difficulty of test items. Leong (2006) understood difficulty levels as an appropriation of knowledge elements

and proposed the following frame for locating item difficulty comprising: content difficulty, task difficulty, stimulus difficulty and expected response difficulty.

According to Leong (ibid), in the assessment of knowledge, the difficulty of a test item resides in the various elements of knowledge such as facts, concepts, principles and procedures. These knowledge elements may be basic, appropriate or advanced. Basic knowledge elements are those that learners have met at lower levels according to the curriculum. Thus, they are very likely to be familiar to candidates because of prior experiences. Advanced knowledge elements are usually those that will be covered more adequately at advanced levels and hence are peripheral to the core curriculum, and learners may not have had sufficient opportunity to learn. These knowledge elements are likely to be difficult for most of the learners. Knowledge elements at the appropriate level are those that are central to the core curriculum within the focal grade. Unlike IRT, Leong's approach to item difficulty is not based on data on empirical learner performance. Rather, items are viewed as easy or difficult for learners in relation to the grade-based curricular specification context. Overall, items that test knowledge elements at the appropriate grade level are viewed as likely to be moderately difficult to learners (Leong, 2006, p.3).

Leong (2006) further suggests that content difficulty may also be varied by changing the number of knowledge elements assessed. Generally, the difficulty of an item increases with the number of knowledge elements assessed. Test items that assess learners on two or more knowledge elements are generally more difficult than test items that assess learners on a single knowledge element. The difficulty of a test item may be further increased by assessing learners on a combination of knowledge elements that are seldom combined (Ahmed, Pollitt, Crisp, & Sweiry, 2003).

These ideas of Leong (2006) offer an expanded understanding for test items being classified as easy, moderate or difficult linked to a grade as described by de Lange (1999). Linked to a specified curriculum, Leong's approach can be applied to both SNA and SBA test analysis. Other elements of Leong's (2006) difficulty frame: task difficulty, stimulus difficulty and expected response difficulty are closely linked to ideas on cognitive demand, language demand and item format which are also discussed as categories in this chapter. Applying Leong's (2006) difficulty frame, an

item based on content drawn from earlier than Grade 6 would be regarded as easy; centrally grade 6-curriculum items as moderate; and multi-step, multiple knowledge dimensions of grade 6 and beyond as difficult. In this study, Leong's elements are used in this way to analyse difficulty of test items of SNA and SBA.

2.5.3 Cognitive demand

Literature viewpoints suggest that cognitive demand can be closely associated to concepts such as cognitive processing ability (Anderson, 2005), depth of knowledge indicators (Webb, 2002) and proficiency standards (expected level of competence) (Postlethwaite & Kellaghan, 2008). Since these aspects of cognitive demand link closely to aspects of difficulty level (discussed earlier), cognitive demand should be appropriately distinguished from difficulty level. Two important differences are clarified below.

Firstly, there is not always a direct correlation between item difficulty and cognitive demand. Statistical procedures can be carried out on specific test instruments and test items to determine correlations between item difficulty and cognitive levels (Ogomaka, 2013). Historically, de Lange (1999) has argued that there is a positive correlation between difficulty level and cognitive demand, implying that questions of high cognitive demand are more difficult to solve and those of low cognitive demand are easier and take up less time to solve. More recent views indicate that this may not always be the case. Researchers like the committee of reviewers working on the New England Common Assessment Program (NECAP), argue that the two terms, difficulty level and cognitive demand, are not synonymous (NECAP, 2013). As an example, "solving a multi-step linear equation with variables on both sides may be a difficult task for middle school students; however, the task can be solved by applying a standard procedure making the task of low complexity" (NECAP, 2013, p.1). The above example illustrates that a test item can be regarded as difficult but with a low cognitive demand.

Secondly, there is a direct link to the learner. Whereas difficulty level deals with the appropriation of knowledge elements (Leong, 2006) cognitive demand refers to the cognitive processing characteristics of assessment items: in other words the

cognitive processing capacity required of a learner in order for him or her to successfully answer the assessment item. Difficulty levels deal more with the complexity of item characteristics of which cognitive demand can be a contributing descriptor. It is possible to determine difficulty level outside of learner actions (excluding IRT approaches) if a classification frame is used to identify and appropriate specific knowledge elements (Leong, 2006). Cognitive demand deals directly with the complexity of thinking levels required from a learner and the philosophy has been to structure these thinking levels to show a progressive increase in the depth of knowledge (Webb, 2002). This progression is related to the number and strength of connections of concepts and procedures that a student needs to make to produce a response (NECAP, 2013).

Cognitive demand is an assessment feature that is prominently referred to in the literature of both SNA and SBA. Several approaches towards understanding cognitive demand can be seen in both contexts. In SNA literature, the general trend has been to view cognitive demand in some order (either hierarchical or non-hierarchical list) of expected mathematical competencies a learner must display towards arriving at a solution. While this might involve a combination of several cognitive processing steps, the descriptors of cognitive demand are matched to an item according to the main competency involved. For test designers to accurately define the mathematical competency required in a test item, they must have an understanding of the cognitive demand evident in the task. In the SNA studies listed earlier (e.g. TIMSS), cognitive demand is listed in three domains ranging hierarchically from knowing, applying and reasoning. In South Africa, the subject assessment guidelines for the national senior certificate (NSC), public exit examination after 12 years of schooling, splits the reasoning domain further into complex procedures and non-routine problem solving (DoE, 2009b).

Postlethwaite and Kellaghan (2008) also pointed out that it is common practice among international assessment agencies to match cognitive demand of test items according to levels of proficiency standards or hierarchical achievement levels. In so doing, this produces competency lists associated with hierarchical cognitive demand levels. An example of this approach for mathematics was used in the Laos 2006 Grade 5 National Assessment Survey (Sisouk & Postlethwaite, 2007). In the Laos

survey, six hierarchical levels are used (see table 2.9) to define skills of learners. This implies that students at Level 1 are unlikely to be able to answer correctly items associated with the skills shown at higher levels. Those at Level 2 would be likely to be able to do what was required at Levels 1 and 2, but not at Levels 3 to 6.

Table 2.9: Mathematics skill levels from the Laos Assessment survey

Level 1: May recognise and classify basic shapes. Familiarity with numbers described in word and numeric form. Understanding of place value for whole numbers.
Level 2: Emerging ability to perform single-step arithmetic operations including addition, subtraction, multiplication, and division. Recognises fractions in both numeric and visual representations. Some concept of symmetry emerging.
Level 3: Emergence of arithmetic applied to problem solving. Multiple-step arithmetic operations. Understanding place value for decimals. May read a value from a simple bar graph. Familiarity with inequalities and ability to order decimal numbers by magnitude. Developing understanding of proportional fractions. Basic conversion of linear units such as length, weight, and time.
Level 4: Developing the ability to solve word problems requiring a fraction or percentage operation. Developing more sophisticated arithmetic including BODMAS, long division, and multiplication with decimals. Conversion between various units of weight, time, and volume. Deals with elementary spatial problems involving 2-dimensional displacement.
Level 5: Beginning to combine and summarise multiple pieces of information from charts; has developed an understanding of spatial concepts such as rotation and reflection; able to convert units for weight, time, area, and volume.
Level 6: Use of rules and symmetry to solve geometric and measurement problems. Strong command of unit conversion for everyday measures. Capacity to solve word problems using a range of appropriate arithmetic operations.

Extending the idea of levels of proficiency, the SACMEQ designers linked competency to a 8-level scale of mathematical proficiency. They defined each of the eight competency (or skill) levels with a descriptive account of the skills that must be acquired for pupils to move from one level of competence to a higher level (Moloi and Chetty, 2011). This “skills audit” for the mathematics tests resulted in the identification of eight hierarchical levels of competence for each test (Level 1 being the lowest, and Level 8 being the highest). The SACMEQ competency levels for mathematics are hierarchically listed as: 1) Pre- Numeracy, 2) Emergent Numeracy, 3) Basic Numeracy, 4) Beginning Numeracy, 5) Competent Numeracy, 6)

Mathematically Skilled, 7) Concrete Problem Solving and 8) Abstract Problem Solving.

The literature on other SNA programmes point towards a clustering of levels so that there are fewer categories to make sense of. In the NAEP assessments, the cognitive demand of questions is described on a 3-tier scale consisting of: 1) Basic; 2) Proficient and 3) Advanced. The NAEP achievement levels *Basic*, *Proficient*, and *Advanced* are used to interpret the meaning of the NAEP scales and are indicators used to interpret student performance. These are different to difficulty level descriptions which are based on an appropriation of knowledge elements independent of student performance. *Basic* denotes partial mastery of the knowledge and skills that are fundamental to proficient work at a given grade. *Proficient* implies students reaching this level have demonstrated competency on challenging subject matter. However, *Proficient* is not synonymous with mastery of grade-level performance. *Advanced* signifies mastery and superior performance (de Mello, 2011).

A non-hierarchical list of general mathematical competencies was defined in the OECD framework for the PISA study to include the following: 1) Mathematical thinking, 2) Mathematical argumentation, 3) Modelling, 4) Problem posing and solving, 5) Representation, 6) Symbols and formal language, 7) Communication and 8) Aids and tools (de Lange, 1999). It was further argued by the PISA designers that in order to operationalize these mathematical competencies, it is helpful to organise mathematics skills into three levels. An example of such organization was the National Dutch option of TIMSS (Boertien & de Lange, 1994; Kuiper, Bos, & Plomp, 1997) later adapted in OECD studies (de Lange, 1999). The three levels were listed as:

1) Reproduction, definitions, computations.
2) Connections and integration for problem solving.
3) Mathematisation, mathematical thinking, generalization, and insight.

The SBA literature shows similar views of thinking about cognitive demand but places more emphasis on the cognitive processing aspects than on proficiency

standards Since Bloom's early work on taxonomy levels, many researchers have used various schemas to describe cognitive demand in SBA contexts (Hess, 2006). A few have been listed in the SBA discussion earlier: the LTA of Anderson (2005), Linn and Miller's (2004) table of specifications and de Lange's (1999) assessment pyramid. A common view among these is that the cognitive demand of assessment tasks at school level should be structured according to hierarchical levels. This view was also considered by Norman Webb (2002) in his 'Depth of Knowledge (DOK)' classification system which is based on four progressive levels of thinking. A brief description of Webb's (2002) classification system is indicated in table 2.10 below.

Table 2.10: Webb's DOK classification system

Level	DOK	Descriptor
1	Recall	Recall or recognition of fact, information, concept or procedure.
2	Basic application of skill/concept	Use of information, conceptual knowledge, follows or selects appropriate procedures, routine problems.
3	Strategic thinking	Requires reasoning, developing a plan or a sequence of steps to approach a problem, abstract and complex.
4	Extended thinking	An investigation or application to real world; requires time to research, think and process multiple conditions of the problem or task; non-routine manipulations.

In the above table the DOK levels name 4 different ways learners interact with mathematical content. Each level is dependent upon how deeply learners need to understand the content in order to respond (Hess, 2006). Hess (2006) points out that Webb's (2002) DOK levels are being increasingly used among state departments and schools in the USA to develop performance assessments to demonstrate learning.

Stein et al (2000), who have done extensive research in the area of characterising the cognitive demands of mathematical tasks in SBA, also use four categories (i.e., memorisation, procedures without connections, procedures with connections, and doing mathematics) for differentiating tasks. The authors suggest that teachers use these four levels of cognitive demand when considering aspects of their classroom practices. Their standpoint is that it is important to know the potential of a task so

that it can be appropriately mapped to student's thinking and ensure balance across low level and high level tasks. Tasks in the first two categories require students to perform a memorized procedure in a routine manner and are referred to as low-level tasks. Tasks in the 3rd and 4th categories that demand engagement with concepts and that stimulate students to make purposeful connections with meanings or between relevant mathematical ideas are considered high-level tasks (Stein et al, 2000).

The DBE's (2011) consideration of cognitive levels in their Curriculum and Assessment Policy Statements (CAPS) are listed as: knowledge, routine procedures, complex procedures and problem solving. For each of these cognitive levels, a description of skills to be demonstrated by the learner is listed. The descriptors are similar to the ideas listed by Stein et al (2000) and Webb (2002) above. The guidance to school teachers was that their forms of assessment should be appropriate to the age and the cognitive level of the learner. There is also advice that questions in tests and examinations in particular 'should be carefully spread to cater for different cognitive levels of learners' (DBE, 2011c, p.294). The highest weighting of questions should be on routine procedures (±45%) and the lowest weighting should be on problem solving (±10%) suggesting a hierarchical structure of the levels in terms of cognitive demand, as noted by de Lange (1999).

The various schemas outlined in this section highlight the range of possibilities for an analysis of cognitive demand on standardised tests (e.g. ANA) as well as the SBA of teachers in this study. The proficiency standards of SNA studies provided valuable information on a hierarchy of skills, and within the SBA literature there is synergy among researchers on ideas associated with levels of cognitive processing and depth of knowledge to explore and unpack cognitive demand. In this study, the ideas of Webb, Anderson and Stein were considered generally aligned to aspects of cognitive demand found in the TIMSS framework and to the cognitive level descriptions put out by the DBE (2011c) in their Curriculum and Assessment Policy Statements (CAPS). Put together, these ideas informed the following order of hierarchical levels for looking at cognitive demand: 1) Knowing basic facts, 2) Applying routine procedures, 3) Applying complex procedures and 4) Solving non-routine problems. Detailed descriptors for these are provided in the methodology

chapter, with these cognitive levels forming a framework for evaluating the cognitive complexity of assessment items (what students are expected to know and be able to do) in this study across SNA and SBA.

2.5.4 Language demand

Although language demand is not listed as an explicit feature of the SNA studies, Leong (2006) argues that considerations of language are an implicit assessment feature that often contributes to the understanding of difficulty level and cognitive demand. By looking at the language demand of questions one can gain valuable insights into the amount and nature of text used by teachers in SBA and test designers in SNA and/or into the kinds of contextualisations selected for their assessment tasks. Language demand provides an indication of how contexts and words are used to assess mathematical skills.

Leong (2006) referred to the difficulty that learners face when they attempt to comprehend the words and phrases in a test item and the information that accompanies the item (e.g., diagrams, tables and graphs) as ‘stimulus difficulty’. He argued that test items can be differentiated on language on the basis of those that contain words and phrases that require only simple comprehension (e.g. basic operations in mathematics) and those that require more technical comprehension (e.g. solving word problems). The language demand can also be influenced by the manner in which information is packaged in a test item. Test items that contain information that is tailored to an expected response (i.e., no irrelevant information in the test item) are generally easier than test items that require learners to select relevant information or unpack a large amount of information in a given context. Using the ideas of Leong, mathematics items can be considered in terms of the amount of text the learner has to unpack.

The inclusion of contexts in “word problems” in assessment tasks usually increases linguistic complexity as they are more words to interpret and decipher. Contexts can play a major role as vehicles for assessing insight, understanding, and concepts. A variety of contexts is needed, as well as a range of roles for the contexts (de Lange, 1999). They can be “real” using factual or familiar knowledge of learners (e.g. actual

names of streets) or they can be imagined where the context used makes use of fictitious knowledge in the stimulus material of questions. A variety of contexts is often advocated in order to minimise the chance of featuring issues and phenomena that are culturally biased or inaccessible to the learner. Meyer (2001) pointed out the roles of context as: (a) to motivate, (b) for application, (c) as a source of mathematics, (d) as a source of solution strategies and (e) as an anchor for student understanding.

Anderson (2005) argues that the three major components in the construct of a question are the introductory material, the stem and the response. The introductory material can take on many forms (e.g. pictures, written or real objects. The stem takes the form of a statement or directive (e.g. Prove this). Finally the expected response can be short or long (extended). Not all assessment tasks include all three components. He notes the need for careful balance between text decoding demands and authenticity demands. Reducing the words used in test items to a bare minimum can reduce translation costs, but it also typically decontextualises the item, thereby making it less authentic. This body of work suggests that it would be useful for test frameworks for SNA and SBA to clarify nature of language demands to be used in the national assessment tasks and also in internal teacher tests.

Historically, language demand has been seen as a key inhibiting factor in South African learner performance in national assessments (Howie *et al*, 2012). This evidence points to the need for test designers to be aware of stimulus difficulty in the construction of mathematical questions. She argued that standardised tests should minimise language bias. Thus, learners should not be penalised or advantaged by life experiences that are not relevant to the knowledge, skills and understandings that the test is intended to assess. Further, the stimulus material should not breach ethical, cultural or other sensitivities. de Lange (1999) pointed out that language contexts used in tests are either “real world” with words, names and places that learners are familiar with or they are imagined. Imagined contexts are authentic but make use of fictitious names and life situations. Language instructions should follow central issues in the stimulus, rather than requiring peripheral knowledge. An item is considered fair when it tells learners what they are required to do; it stands alone and does not depend on understanding from a previous item (Howie, 2002).

Reading to acquire and use information is regarded as a critical skill in the primary school, and is the purpose of international comparative reading literacy studies such as the Progress in International Reading and Literacy Study (PIRLS) (Howie et al, 2012). The 2004 SE study (DBE, 2005a) indicated a positive correlation between reading skills and mathematical performance, suggesting relationships between language demand and learners' responses to test items in mathematics tests is significant. Therefore in a standardised assessment such as the ANA where learners are required to interpret information on their own, the nature and amount of text in test items can be a critical factor in test design.

In this study, the language demand (LD) of an item referred to the amount and nature of text in the instructions and stimulus material and three categories was adopted: no text questions; low text questions and high text questions. In the methodology chapter, descriptors based on the work of Anderson (2005), Leong (2006), and Howie (2002) exemplify these categories examining the amount and nature of text included in a test item.

2.5.6 Item format

Finally in both SNA and SBA, the format of questions is a relevant indicator to establish the balance of assessment tasks in terms of offering learners varied and multiple opportunities to present their mathematical skills. When engaging in classroom assessment, the teacher is confronted with many tasks, choices, and dilemmas. Anderson's (2005) taxonomy table and Linn and Miller's (2004) concept of a content specifications table suggest some of the decisions teachers make in choosing a desired format of a task to assess specific learning outcomes in mathematics.

In paper-and-pen assessments, traditionally associated with SNA, learners respond to a series of questions or prompts. Their written or drawn responses are used as evidence of their level of knowledge, competence, or understanding. There are generally four basic item formats, or ways that learners can show their responses, with the first two more widely used in mathematics test design (Anderson and

Morgan, 2008):

a) Multiple choice
b) Closed constructed response
c) Open-ended short response
d) Essay or extended response

Multiple-choice items require learners to select one of several (usually four) options. Multiple-choice items usually have one unequivocally "correct" option and several plausible but incorrect options (referred to as distractors) (Anderson and Morgan, 2008).

Table 2.11: Example of a Multiple-Choice Item

Marcel has 3 flowers. Dad gives him 2 more flowers. How many flowers does Marcel have in all?
A 2
B 3
C 4
D 5 (correct answer)

Source: Adapted from Anderson and Morgan (2008).

In constructing an achievement test to fit a desired goal, the test maker has a variety of item types from which to choose. It will come as no surprise that the multiple-choice format seems to be the "best" format if we simply judge by its popularity in SNA studies such as TIMSS and SACMEQ. Multiple-choice, true-false, and matching items all belong to the same category: selection-type items. Officially, they are popular in SNA designs because they are an item type that is scored objectively (de Lange, 1999). They are low-cost in terms of time and labour-efficient in terms of marking for large-scale assessment.

Closed constructed response items have one correct answer that the student generates. Minor variations in the way the answer is shown are usually acceptable. Learners may be required to write a few words, show calculation steps, underline a word or number in a text or table, draw a line on a grid, or indicate an area of a diagram. In this sense, they may also be referred to as closed short answer (CSA).

In this format learners may also be required to select several options that meet certain criteria or to match a series of pairs of sentences or diagrams. CSA types of items are evident in the NAEP Mathematics Test (1990–2000) for Grade 4 and IEA assessments such as TIMSS tests (e.g. TIMSS 2003 Grade 4 Mathematics Test (Anderson and Morgan, 2008). An example of a CSA item format is provided in **Table 2.12**.

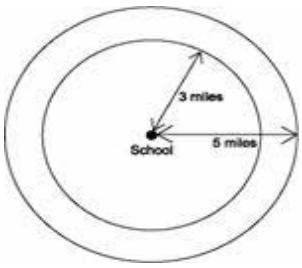
Table 2.12: Example of a Closed Constructed-Response Item

Here is a number sentence: 2,000 + ☐ + 30 + 9 = 2,739 What number goes where the ☐ is to make the sentence true? Answer: _____

Source: IEA 1998, sample item.

Different item formats can be combined in the same test. For example, a test may consist of some multiple-choice items, some closed constructed-response items, and some open-ended short response items. Open ended and extended response open questions give learners the opportunity to get involved in a context with one or more open questions of relatively complex nature, where the learner's choice of which strategy to follow is not clear in advance. Extended response differs from open questions in that the students are expected to explain their reasoning process as part of their answer (see **Table 2.13** for an example).

Table 2.13: Example of an Extended Response Item

"Martin is living three miles from school and Alice five miles. How far apart are Martin and Alice living from each other?" The answer "they live two miles apart" is just part of the correct answer. Students may make a drawing to explain their reasoning:  Martin and Alice could live 8 miles from each other, or 2 miles or any number in between.

Source: de Lange (1999): framework for classroom assessment

In this study, the two types of item format (IF) considered for analysis and discussion are multiple choice questions (MCQ) and closed short answer type (CSA). In the methodology chapter, descriptors based on the work of Anderson and Morgan (2008) and de Lange (1999) are provided for examining item format.

2.6 An interpretive map for analysing SNA and SBA

From the literature review, an interpretive map of advocated design features, considered as 'rules' relating to motives and goals can be constructed. Policy documentation on SNA and SBA provides an indication of collective goals with broad or specific mandates; individual goals linked to teacher roles; and operational goals that direct action in the design of instruments. I use Morgan's (2000) assessment discourse framework to understand advocated rules and locate them within specific strands, namely, psychological, curriculum implementation and curriculum standards.

The literature review indicates that when a multiplicity of goals is evident, the nature of SNA straddles elements across different discourses. Multiple goals can be a source of shifts between assessment discourses and generate tensions. When disaggregated, cross-national and regional assessment SNA designs mainly focus on alignment to curriculum expectations. In some countries (like South Africa) SNA designs attempt to gather information on curriculum implementation and the psychological needs of the learner. In this scenario, there are often trade-offs that compromise the effectiveness of the design to achieve any of the multiple goals. The goals of SBA suggest an alignment towards curriculum implementation with concerns for the psychological needs of the learner also at play. With an understanding of goals influencing actions, a more nuanced look at rules governing assessment design of SNA and SBA becomes possible.

Expanding on Morgan's assessment discourse, the second important consideration is to map the relations between rules, goals and tools and understand the effect such relations have on assessment activity. Literature suggests that a useful way to understand these relations is to consider it in the context of an activity system. In this study a discussion is taken up on the structure and key components of SNA and

SBA as activity systems to further understand the influences of motives and goals on design rules and the alignment of advocated rules to implemented artefacts (tools). Literature on activity systems linked to the nature of advocated rules, multiplicity of goals and assessment artefacts are detailed in Chapter 3 when I discuss activity theory (AT). Using AT concepts, relationships between the design rules of SNA and SBA and what was found in the assessment artefacts can be analysed. In SNA, the literature suggests that rules that influence the design of artefacts can be seen in a spectrum of established cross-national and regional assessments (e.g. TIMSS) as well as local national assessments (e.g. Systemic Evaluations in South Africa). In SBA, the rules may be explicitly seen in design frameworks, or need to be inferred more implicitly from the assessment tools used by teachers. In this study, the ANA (from 2008-2010) and the SBA of the three teachers were considered as the key assessment artefacts to explore.

The reviewed literature suggests that the following design rules are applicable across SNA and SBA: 1) coverage across the content of mathematics, 2) level of difficulty in questions, 3) cognitive demand, 4) language demand and 5) the item format. Each of these aspects has an extensive literature base in mathematics education and assessment across SNA and SBA. When looking at their application in specific assessment tools, these aspects provide useful item characteristics that can be summarised into item maps and compared. In the research design chapter, these aspects are clustered into two themes: mathematical coverage and the range and scope of questions. In the SBA context, a third theme, referred to as the analysis of common topics, was added to further explore the how teachers dealt with design features on individual topics.

The literature in SNA is emphatic that, whatever the frame used, there should be questions at all levels of thinking, of varying degrees of difficulty, and in all content domains (de Lange, 1999). In SBA, similar principles apply but the content coverage may be restricted to a single domain or topic in mathematics. The summary table below provides a literature map of the categories that were used for analyzing SNA and SBA Mathematics assessments, referenced to the literature discussed in this chapter.

Table 2.14: A literature map for analysing test item categories in SNA and SBA

Categories	Sub-categories	Source
Coverage:	Number	DBE (2002), SACMEQ (2007) and TIMSS (2011)
	Patterns	DBE (2002);
	Shape and Space	DBE (2002) and SACMEQ (2007
	Measurement	DBE (2002) and SACMEQ (2007)
	Data	DBE (2002), SACMEQ (2007), and TIMSS (2011)
Difficulty level	Easy (E)	Leong (2006), de Lange (1999)
	Moderate (M)	Leong (2006)
	Difficult (D)	Leong (2006), de Lange (1999)
Cognitive Demand	Knowing basic facts (K)	TIMSS (2011), Stein et al (2000) (Anderson (2005), Linn and Miller (2004), de Lange (1999), Webb (2002), DBE (2011)
	Applying routine procedures (A)	TIMSS (2011), Stein et al (2000) (Anderson (2005), Linn and Miller (2004), de Lange (1999), Webb (2002), DBE (2011)
	Using complex procedures (C)	TIMSS (2011), Stein et al (2000) (Anderson (2005), Linn and Miller (2004), de Lange (1999), Webb (2002), DBE (2011)
	Solving non-routine problems (N)	TIMSS (2011), Stein et al (2000) (Anderson (2005), Linn and Miller (2004), de Lange (1999), Webb (2002), DBE (2011)
Language Demand	High text (word problems) (HT)	Leong (2006), DBE (2002), de Lange (1999)
	Low text (instruction/s is/are short mathematical verb/s) (LT)	Howie (2002), Leong (2006)
	No instructional text (Computational) (NT)	Leong (2006)
Item Format	Multiple choice questions (MCQ)	Anderson and Morgan (2008), de Lange (1999)
	Closed short answer response questions (SR)	Anderson and Morgan (2008), de Lange (1999)

Drawing from the above frame, item maps were developed and used for critically

analysing and interpreting assessment artefacts for both SNA (with the ANA tests as the key artefacts) and the SBA tasks of selected teachers. These are explained further in the methodology chapter.

2.7 Conclusion

The aim of this chapter was to present a literature review of key underlying concepts linked to SNA and SBA as observed in international literature and in the South African assessment landscape. In order to better understand work presented later, I have highlighted the sources of concepts and connections between them. The issues raised here foreshadow the discussion in later chapters. A standpoint going forward is that the literature on SNA suggests that national assessment discourses can exert significant influence on SBA practices. In this study I view this influence particularly in terms of 'rules' that circumscribe action, and the nature of activity systems and elements found in assessment artefacts. Further theoretical detail that provides a handle of this view is provided in Chapter 3, where activity theory is detailed. This is presented next.

3. Theoretical framework

3.1 Introduction

In this chapter a theoretical interpretation of design rules in SNA and SBA and the nature of their interaction is offered. Firstly SNA and SBA as activity systems must be understood in terms of their individual structural make-up. The literature viewpoints suggest that advocated rules give shape to the kinds of assessment artefacts (tools) found within SNA and SBA. Therefore in this chapter, theoretical relationships between goals, rules and tools as part of an activity system are explained. Secondly, the literature review suggests that within the make-up of SNA and SBA there are tensions and contradictions between what are advocated as rules and what are found in the assessment artefacts. In this chapter, I propose that these tensions are related to a broader range of goals and motives at a policy level that exert significant influence on assessment design rules. The goals and motives of SNA are implicated within SBA rules but the SBA system also has other more local goals so there is also potential for tensions and contradictions. Therefore the nature of tensions occurring within and across activity systems must be explained.

The argument made in this chapter is that an exploration of SNA and SBA is better understood not just as two types of assessment that historically co-exist in the assessment landscape but as two interacting systems of educational assessment structurally defined by their individual subcomponents and by the nature of activities occurring within and across systems. The purpose of this chapter is then to link this argument to a theoretical framework that provides a handle on these matters. In this chapter, activity theory (AT) is proposed as a suitable theoretical lens for understanding SNA and SBA as interacting assessment systems.

AT is proposed as being better equipped than other socio-cultural perspectives to explain the structure and alignment of SBA and SNA as activity systems, and offers a way to understand tensions within and between them. The issue of context in relation to human activity is foregrounded. In this research study, there was a context of assessment in relation to educational activity in mathematics that needed to be understood and explained. From an AT perspective, context is not simply a container or a ‘situationally’ created experiential space but is an entire activity system,

integrating the participant, goals, tools (and communities, their rules and divisions of labour) into a unified whole (Engeström, 1993).

The chapter begins with an overview of AT outlining its historical development over three generations. The third generation of AT (Engeström, 1987, 1993, 1999; Leont'ev, 1974, 1981) is then explored as the preferred theoretical and methodological AT application for describing and analysing the assessment activity systems under investigation. The ideas of Vygotsky (1978) on mediated action and Leont'ev (1981) on the hierarchical structure of activity linked to motives, goals, and operational conditions are considered within Engeström's model.

The generic AT elements of Engeström's model are highlighted. Thereafter the application of these elements to SNA and SBA are flagged, with these elements discussed in greater detail in the analysis chapters that follow later in the study. The central tenets of AT that I selected to feature in the analysis of SNA and SBA are explained.

An AT perspective on identifying tensions within and across systems is provided which forms the basis for understanding contradictions within and across SBA and SNA identified later in the analysis chapters. Evidence on the use of AT in broad educational contexts and specific mathematics education studies punctuates arguments on the suitability of AT in this study. In particular, reference is made to ways in which the mathematics educational community has operationalised the activity system as a unit of analysis. The chapter concludes with an outline of implications for using AT in this study.

3.2 An overview of activity theory

According to van Oers (2001, p. 71), the concept of activity, which refers to "any motivated and object-oriented human enterprise, having its roots in cultural history, and depending for its actual occurrence on specific goal-oriented actions", is central to a sociocultural approach. Activity, as synthesized by Daniels (2001, pp. 84-86) with reference to Davydov, Leont'ev, and Engeström, has a developmental function, is characterized by constant transformation and change, is guided by a motive, and is a collective and systemic formation that has a complex mediational structure

(Jaworski & Petoni, 2009). Activity in this sense is considered to be both socially and culturally determined (Westberry, 2009; Kaptelinin, Nardi, & Macaulay, 1999).

As this study involved an investigation of the nature and discrepancies in assessment activities in mathematics at a school level influenced by national policy texts, AT afforded me the opportunity to explore the influence of the ANA on the tools used by individual teachers' SBA within a community of practice.

3.2.1 Historical development - three generations of activity theory

Although many writers have contributed to develop AT concepts, the focus here is on the seminal work of three main authors: the Soviet psychologists Vygotsky and Leont'ev, and the Finnish researcher Engeström.

The history of AT can be represented by three distinct generations (Engeström, 2001). The first generation was characterized by work on mediation (Vygotsky, 1978) whose ideas formed the classical building blocks from which other contributions stemmed. The second expanded the unit of analysis to include the social context (Leont'ev, 1981), and the third generation expanded the single context analysis to include two or more activity systems (Engeström, 2001).

3.2.1.1 First Generation

The first generation, centered on Vygotsky's early work, elaborated the idea of mediation. This idea was crystallized in Vygotsky's (1978, p. 40) famous triangular model (Figure 3.1) in which he expressed his idea of cultural mediation of action in terms of the triad of subject, object, and mediating artifact (Engeström, 2001). Vygotsky's cultural-historical psychology is regarded as the first generation of activity theory or Cultural Historical Activity Theory (CHAT) and later became the basis for modern applications.

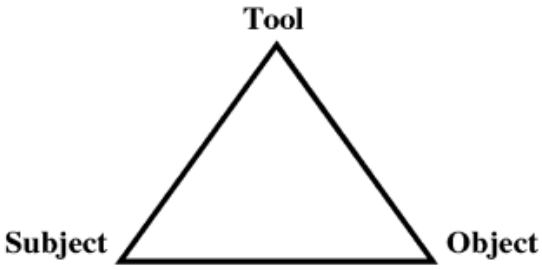


Figure 3.1: First generation activity theory: Vygotsky's (1978) model of mediated action.

Proponents of AT (Leont'ev, 1978; Engeström, 1987; Wertsch, 1991) based their work on the Russian cultural-historical school, extending Vygotsky's (1978) model of tool mediated object-directed action (figure 4) to collective activity (Venkatakrishnan, 2004). In Vygotsky's early work, the unit of analysis was object-oriented action mediated by cultural tools but did not integrate the collective social interaction of a community into the triangular model of action (Engeström and Miettinen, 2003). But Vygotsky had argued that emphasis must be placed on the 'collective societal' rather than the 'individualistic dimensions' of practice (Roth & Tobin, 2002). Thus, the individual could no longer be understood without his or her cultural means; and society could no longer be understood without the agency of individuals who used and produced artefacts (Engeström, 2001).

The limitation of the first generation was that the unit of analysis remained individually focused and as Barab et al (2004) noted, Vygotsky did not fully develop the concept of activity in his brief lifetime; thus, the task of articulating the nature of activity fell to his colleague, Leont'ev (1981), resulting in the second generation of activity theory.

3.2.1.2 Second generation

The work done by Leont'ev was regarded as the second generation of AT (Westberry, 2009) and others have argued that he was one of the first to articulate the concept of activity and its role within a larger activity system (Barab et al, 2002b; Jurdak, 2009). Some argue that AT is therefore based on the collective heritage of the founders, in particular the *troika* of Vygotsky, Luria, and Leont'ev (Sannino, Daniels, & Gutiérrez, 2009). In building up his structure of the human activity system,

Leont'ev, in his initial work (1978), conceptualized the distinction between collective activity and individual action by introducing a hierarchical 3-level model of activity (see Figure 3.2 below). The uppermost model of collective activity is based on an object-related motive; the middle level of individual action is based on a goal; and the bottom level of operations is based on tools and conditions of actions (Engeström & Miettinen, 2003).

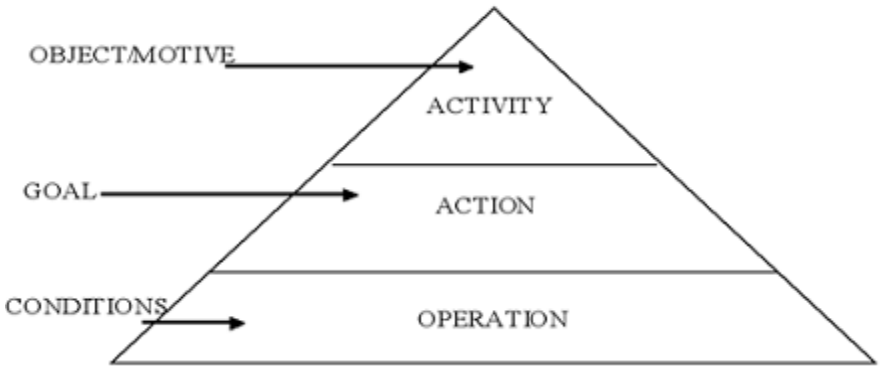


Figure 3.2: Leont'ev's hierarchical structure of activity.

In terms of activity, each level is associated with a special type of unit. The first most general level is associated with the unit of activity that deals with a collective motive or system driven activities such as work, play, or systemic assessment. The second level of activity focuses on the unit of more individual goal driven action. The third level of analysis is associated with the unit of operation or the conditions under which the action is carried out. Operations help actualize the general goal to make it more concrete (Jurdak, 2009). Leont'ev (1981) argues for viewing operational and individual actions within wider systems of collective activities.

However, not all proponents of AT were satisfied that the complete picture of human activity was given by Leont'ev. Michael Cole (1988) was one of the first to clearly point out the insensitivity of the second generation AT toward cultural diversity. As AT went international, questions of diversity and dialogue between different traditions or perspectives became increasingly serious challenges. It was these challenges that the third generation of AT set out to deal with (Engeström, 2001).

3.2.1.3 Third Generation

Whereas Vygotsky began the process of moving the locus of cognition and knowing more generally outside of the individual mind, and Leont'ev refined the emphasis on actions as part of larger activities, Engeström (1987) expanded on Leont'ev's model to better understand cultural diversity and the role of historically formed mediating artifacts. Kaptelinin, Nardi, and Macaulay (1999) remarked that Engeström proposed a scheme of activity different from that of Leont'ev; it contained three interacting entities: the individual, the object and the community; instead of the two components: the individual and the object, indicated in Leont'ev's original scheme. Kuutti (in Engeström et al, 2003) described the structure of an activity as: "a form of doing directed to an object, and activities are distinguished from each other according to their objects". Blackler (in Sannino, Daniels, & Gutiérrez, 2009, p.29) points out that Engeström's most significant contribution to activity theory was his suggestion, "that rather than the socially mediated individual being taken as the basic unit of analysis, the historically located activity system should be the fundamental unit".

Engeström (1987) listed the minimum elements to include in a model (Figure 3.3 below) referred to as the expanded mediation triangle (EMT). He listed these elements as: object, subject, mediating artefacts (signs and tools), rules, community and division of labour.

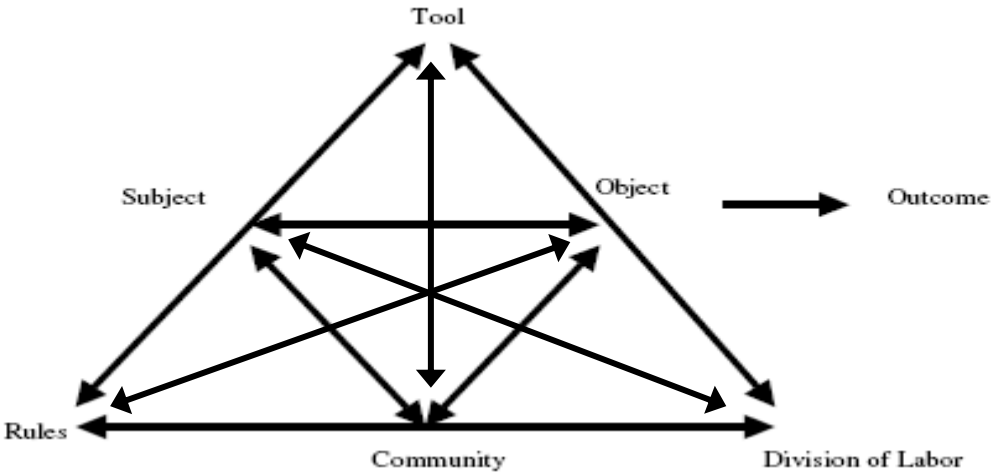


Figure 3.3: Engeström's (1987) expanded mediation triangle (EMT)

AT is most often used to describe activities in a socio-technical system through these

six interrelated elements (Bryant *et al*, 2005). These elements are described as:

1. Tools (instruments) or tool mediation - the artefacts (or concepts) used by actors in the system. Tools influence actor-structure interactions, they change with accumulating experience. In addition to physical shape, the knowledge also evolves. Tools are influenced by culture, and their use is a key means for the accumulation and transmission of social knowledge. Tools influence both the agents and the structure. Verenikina *et al* (1998) argued that, tools are "social objects with certain modes of operation developed socially in the course of labour and are only possible because they correspond to the objectives of a practical action."
2. Subject - actors engaged in the activities. Lektorsky in Sannino, Daniels, & Gutiérrez, (2009) argued that “activity has its bearer”. If it is a collective activity, there is a collective subject. If it is activity of an individual, there is an individual subject. In both cases, a subject is not something that generates activity from the outside. Collective activity and mediation are crucial for understanding an individual subject.
3. Goal - also known as the object of the activity system. Goals refer to items that are considered objective but may be linked to social and cultural properties. Goals can be raw materials, conceptual understandings, or even problem spaces, “at which the activity is directed and which is molded or transformed into outcomes with the help of physical and symbolic, external and internal tools” (Barab *et al*, 2002; Engeström, 1993). In this study, the term goals are used instead of object. When there are several goals at play, it may be useful to also consider the underlying motive. Motives then become a collective driver of individual goals. Motives, goals and operations together form a hierarchical structure of activity (Leont’ev, 1981) Motives in this study were seen as the overall goal of assessment in the SNA and SBA contexts.
4. Rules - conventions, guidelines and norms provide the rules for regulating activities in the system. Rules, when explicitly enforced, are an instance of the overt exercise of authority (Sannino, Daniels, & Gutiérrez, 2009). The shape of tools and goals are often directed by what are advocated as rules (an important standpoint in this study).

5. Community or externalization - social context; all actors involved in the activity system. The community of a system refers to those individuals, groups, or both who share the same general objects, and are defined by their division of labour and shared norms and expectations.
6. Division of labour - social strata, hierarchical structure of activity, the division of activities among actors in the system.

The components of activity systems are not static components existing in isolation from each other but are dynamic and continuously interact with the other components through which they define the activity system as a whole. The constituents of activity dynamically change as conditions change (Nardi, 2006b). From an AT perspective, an examination of any phenomenon (e.g. classroom assessment of learners) must consider the dynamic among all these components (Barab *et al*, 2002).

The continuous interaction and mediation of these elements and their internal tensions and contradictions form an activity system that includes both 'historical continuity and local, situated contingency in the analysis' (Engeström & Miettinen, 2003). Mediation occurs between the various components of the activity system with 'other' components (Kuutti, 1996). The top part of Engeström's model represents a goal-directed or purposeful interaction of a subject with an object through the use of tools. In the bottom part of the model, the relationship between community and subject is mediated by rules of behaviour which are explicit and implicit norms and conventions governing social interaction (Westberry, 2009). Also, the relationship between community and object is mediated by the division of labour which is the explicit and implicit organisation of a community as related to the transformation process of the object into the outcome (Issroff & Scanlon, 2002, p. 78).

Engeström's third generation AT model involves the consideration of two interacting activity systems (see Figure 3.4 below) as its unit of analysis, with a "potentially shared object", that is focused on the challenges and possibilities of inter-organisational learning (Engeström, 2001). In this study, the assessment of learners is considered a shared object between the SNA and SBA activity systems.

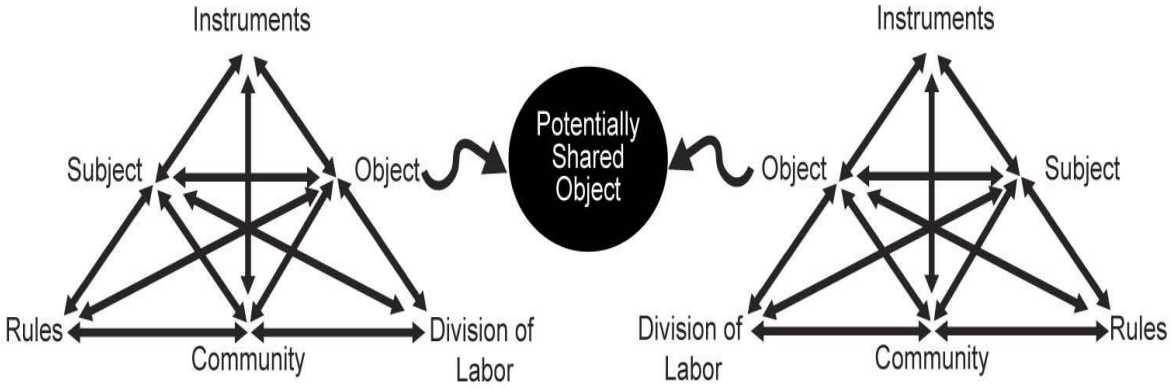


Figure 3.4: Third generation AT model of interacting activity systems

Engeström's 'minimal model of interacting activity systems' has formed the basis of studies within the 'third generation of activity theory' (Engeström, 1987). The focus of research is on interacting activity systems with attention on the mediation of activity within the individual sub-systems and also on interactions across activity systems. In this context, Jurdak (2009) described Engeström's work as the construction of an activity system, developed to describe and account for the collective (as compared to individual) activity systems in broad historical-cultural-social contexts.

Given the above characteristics of three generations of AT, it is the third generation AT that is considered as the principal lens for understanding the nature of SNA and SBA activity systems while the ideas of Leont'ev are used to understand the structure of advocated rules. Using Engeström (1987), I take forward the idea of the activity system rather than the individual as the unit of analysis. Also the model provides a handle to focus on goals, tools and rules of the SNA and SBA activity systems, and secondly, to understand their alignment to each other within a potentially shared and interacting environment. Discussing the SNA and SBA activity systems in terms of Engeström's (1987) model enables a deeper understanding of the two assessment systems under investigation and allows for theoretical explanations about the nature of mediated relationships that exist within and between the various components (such as rules, goals and tools) of the AT systems that impact on the assessment of learners.

The literature reviewed in Chapter 2 indicated several overlaps in advocated design

rules, while noting key differences in their orienting goals. These overlaps and tensions provide a rationale for SNA and SBA activity systems to be considered as interacting as they both share and are shaped by common assessment features and the policy prescripts of SNA further form the rules for the SBA judgments of learners' capabilities in mathematics. In the South African assessment context described in earlier chapters, data on the 'gaps' between SNA and SBA, suggest that the rules emanating from the national policy context are not being applied as intended. Venkatakrishnan (2004) had a similar context in her research on understanding mathematics teacher actions in policy directed contexts in the United Kingdom. She used AT concepts to analyse differences in the take-up of national policy advocated tools with associated goals in two different schools with different histories and goals related to their mathematics teaching.

3.3 Constituting the SNA and SBA activity systems

Conceptually in this study the two assessment systems, SBA and SNA, were considered as units of analysis with the shared object of assessment of mathematics. Applying the ideas of Engeström's (1987) EMT model, I could constitute AT elements in SNA and SBA.

3.3.1 AT elements identified in SNA

The literature discussed in Chapter 2 noted that SNA artefacts (tools) could be developed locally (e.g. the ANA in South Africa) or by international assessment agencies such as the IEA who make available assessment instruments for their TIMSS and PIRLS studies. In this study, the ANA is considered a key artefact of the South African SNA context, which later (in Chapter 5) is referred to as the SA-SNA context. The specific tools considered empirically are the ANA tests from 2008 to 2010, guideline documents and the ANA reports that have been published by the DBE. Broader SNA tools considered in SA-SNA context are the RNCS policy document, the FFL assessment framework and the education sector plan (known as the *Action Plan to 2014: Towards the Realisation of Schooling 2025*). These are the tools under investigation in the SNA activity system and are explained in detail in Chapter 5.

Generally in SNA, the goal of national ministries of education (e.g. the DBE) is to utilise a standardised assessment measure to assess learners' ability levels in a specific disciplinary area (e.g. mathematics). In the SA-SNA context, this goal sits alongside other goals related to getting compliance and alignment in terms of SBA development. In the SNA analysis, I flag the ANA implementation process as having varied goals serving a broad educational community, while at the same time benchmarking the rules of design in the SBA space. These rules, as indicated in the previous chapter, may be driven by psychological needs of the learner, curriculum implementation or curriculum standards (Morgan, 2000).

The subjects of the assessment programme are usually individual test developers (made up of curriculum or assessment specialists) or assessment agencies (e.g. IEA) involved in the development of assessment tools. Some test developers may be teachers considered as experts in their field of work. The rules of SNA are guided by a selection of the specific design features suggested in Chapter 2 on coverage and the range and scope of questions. The rules are generally explicit and informed by an assessment framework or test blueprint (e.g. the SE test framework).

The community involved in SNA programmes may be centrally controlled and just include officials and/or independent agents/reviewers but if the assessment programme is large scale, it is likely at different points of implementation to include politicians and policy makers, test developers, provincial/district officials, principals and school mangers, parents, teachers and learners. The division of labour is then structured according to the various role players involved. Politicians and policy makers provide the assessment mandates and policy goals for SNA programmes. Test developers design the assessment tools according to agreed, specified frameworks. Provincial and district officials provide support and mediation on goals of assessment programmes and may develop supplementary materials to prepare schools and learners for participation and familiarity on assessment tasks. Recently (in 2013), the Free State province in South Africa prepared on-line interactive lessons as additional preparation for their schools participating in the SACMEQ IV study as well as the ANA tests, adding to the evidence base that a key goal in the SA-SNA system is to orient assessment activity in the SA-SBA system towards the

rules of the SNA system:.

> *Regular assessments of high standard are important to track learner progress in all subjects and Grades. Improved learner attainment and retention promotes access to quality education by all learners (FSDOE, 2011)*

In SNA programmes, principals, school managers and teachers are involved in the test administration. In some SNA programmes such as TIMSS and PIRLS, independent administrators may be hired to carry out this task. Most SNA programmes are carried out on a sample of learners and may involve parent and teacher questionnaires. I indicated in Chapter 2 that the ANA is one of few national assessments, worldwide, to carry out census type assessments involving all learners at the primary school level. The division of labour in SNA systems may further be linked to national goals and district officials may be further required to monitor compliance and alignment in terms of SBA development of teachers.

In Chapter 5, a more detailed account of elements within the SA-SNA AT system is provided. At this stage, I flag again that broader research evidence indicated in Chapter 1 suggested that teachers experienced difficulties and tensions in implementing national policy goals calling for compliance with SNA assessment rules. These tensions became aspects to explore in the analysis of the AT elements identified in this study.

3.3.2 AT elements identified in SBA

A significant portion of empirical evidence in this study was drawn from the tools used by teachers in their SBA activities. Tools included all the SBA documentation used by teachers: tests, assessment frameworks, RNCS assessment guidelines, district assessments and associated documentation. The nature and form of these tools provided valuable insights into the design activities of teachers that figured in SBA. AT emphasizes that human activity (such as the design of assessments) is mediated by tools in a broad sense. Tools are created and transformed during the development of the activity itself and carry with them a particular culture and historical remains from their development. So, the use of tools is an accumulation and transmission of social knowledge.

In SBA, the subject is the mathematics teacher whose responsibility is to construct internal assessments to fit the requirements of the local (school) assessment policies and informed by the 'rules' of the curriculum drawn from the SNA policy environment. In the literature, compliance with local assessment goals points towards psychological motives aimed towards monitoring individual progress of learners at different time points in the academic year and their eventual progression and promotion at the end of year.

The community in SBA would generally involve all agents whose roles are to ensure that internal assessments provided to learners are in keeping with policy goals. At the school level, community agents include subject teachers, moderators, the principal, district officials, and parents. The division of labour would reflect various responsibilities in this community. District officials often play an intermediary role (between the national ministry and the school) in this regard and mediate national policies through guideline documents, circulars, memos and face-to-face meetings. Principals direct assessment activities at school in consultation with parents and school management teams. Senior teachers (such as Head of the Mathematics Department) are directly involved in the moderation and quality assurance of assessment tasks. Teachers are responsible for the selection/design of the assessment tasks that they will administer to their learners.

The rules for the internal assessments are usually explicitly stated in some of the SNA policy tools on curriculum and assessment. National assessment policies of ministries sometimes provide details on assessment accountability and alignment. These details can range from the number of assessments a teacher must conduct in a school term, to the type of assessment (e.g. assignments/projects/tests), recording and reporting artefacts, and feedback to parents. Some rules relate to promotion and progression requirements of learners to move to the next grade. The 'high stakes' nature of these rules requires teachers to spend time on internalising them. In this study a key focus was on establishing the extent to which SNA rules found expression as SBA design features that teachers utilised or referred to in their construction of assessment tasks. In South Africa, for example, the RNCS and the FFL milestones provided the broad scope for assessment coverage, with more prescription on sequencing and pacing than the RNCS. While this meant tighter rules

for curriculum coverage dictated by SNA, space for interpretation of SBA assessment rules by teachers remained. According to the South African policy rhetoric, it is the responsibility of the teacher to interpret and understand these rules and then design appropriate assessments to fit the requirements.

> *Tests and examinations are individualized assessment tasks and should be carefully designed to ensure that learners demonstrate their full potential in Mathematics content. The questions should be carefully spread to cater for different cognitive levels of learners (DBE, 2011c, p.294).*

In summary, elements identified in the SNA and SBA activity systems point towards gaps between the individual subject and local realities of the assessment community, making it useful to study both systems through their mediating activities and relationships and the subsequent impact on goals, rules, division of labour and ultimately on the assessment tools. Analysing such relationships implies analysing the different elements firstly on their own and then on their mediating effects within and across systems. Using Engeström's third generation AT system of interacting activities, it was noted that SBA artefacts overlap and contrast with the goals identified in the SBA context inferred from an analysis of national policy documentation, and contrast too with the ANA assessments in the SNA context, viewed as SNA artefacts.

Researchers (Daniels, 2009; Jurdak, 2009; Westberry, 2009; and, Barab et al, 2002) have argued that Engeström's expansive third generation of AT is a set of basic principles that constitute a general conceptual system, rather than a highly predictive theory. It is increasingly impacting on research conducted in collective fields of inquiry such as teaching and learning. The benefit is that AT offers a framework for describing and understanding activities that mediate the relationship between participant and object (Barab et al, 2002) as well as the collective and the object (Jurdak, 2009).

3.4 Looking within and across activity systems: Overlaps, tensions and contradictions

3.4.1 Hierarchical structure of activity

Although a feature associated more with Leont'ev than Engeström, the third generation AT frame does not preclude hierarchical considerations of activity which can be useful in better understanding the presence of advocated rules within the SNA and SBA activity systems. The figure below gives an indication of the hierarchical relationship between motives, goals and conditions in an activity. The highest level is the motive which can be culturally or historically located. An example of a motive would be improving Mathematics test scores of learners at a Grade 6 level. Goals can be seen as the middle level where specific actions are linked upwards to a broader motive and directed downwards towards specific operations. The nature of a goal can be viewed in relation to a collective activity, driven by a motive. Referring back to Morgan's discourse analysis on assessment, there may be some overlap in locating goals according psychological, curriculum implementation and curriculum standard strands. The operations in the bottom level provide the conditions for the goals to translate into meaningful actions.

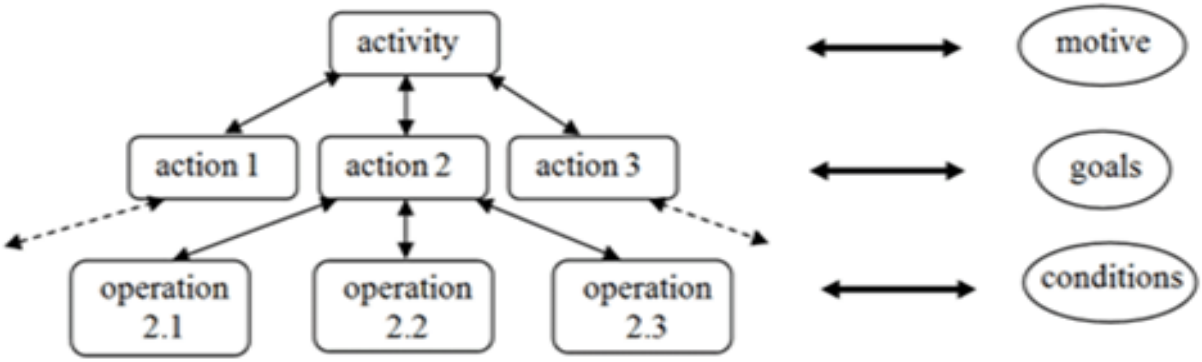


Figure 3.5: Hierarchical structure of activity

This view of activity provided a means to describe and hierarchically organise the policy advocated rules within SNA and SBA contexts. At the highest level would be the motives that relate to the collective policy context and mandates on assessment. This would be followed by goals emanating from the broader political mandate. A typical goal would be the introduction of standardised testing through regular national assessments. An important operation linked to this goal would be the design features stipulated for the creation of national assessment tools that are fair and credible with possibilities for exerting conditions for compliance and alignment with SBA development, i.e. functioning as "rules" for SBA.

To get further insight on these matters an analysis of the policy goals in the SNA

documentation, and of teachers' goals within SBA from interviews, was undertaken. Thereafter the extent to which teacher goals could be seen as influenced by the need for alignment towards state driven assessment goals and individual goals became the focus. Teachers' design features provided a sense of how these goals translated into actual assessment tasks in the SBA programme.

In this study, within the contexts of SNA and SBA, the hierarchical structure of activity was also viewed from two levels: 1) advocated design features informed by state driven policy prescripts and 2) the enacted design features found in the assessment artefacts extrapolated in content coverage and the range and scope of questions. Within the SBA context, the enacted design features also took into account teacher preferences in the assessment of specific topics. Particular attention was given to the extent of contradictions and tensions that existed within the hierarchical structure of activity within these contexts and levels.

3.4.2 Contradictions and tensions

Engeström (1987) suggested that differing goals cause instability and tensions within systems that need to be negotiated. According to Engeström (2001), activity systems are driven by communal motives that are often difficult for individual participants to articulate. He distinguished between communal motives at the level of the collective and individual level actions and argued that activity systems are constantly developing as a result of contradictions and disturbances. Contradictions are best understood as tensions among the components of the activity system (Barab *et al*, 2002) but they can also occur when an individual's goals and tools don't quite align with the activity system motives. As tensions enter the system they become the moving force behind disturbances and innovations and eventually drive the system to change and develop.

Engeström identified four types of contradictions in activity systems (Kaptelinin, Nardi, & Macaulay, 1999) in his research, which can be customised for further research on assessment:

a) Primary contradictions are inner contradictions within each of the nodes of an activity system. For instance, where the assessment tools of a teacher are

designed to include questions covering an extensive spread of cognitive and difficulty levels but then these tools become too costly to produce and distribute to learners, a primary contradiction is seen within the 'tools' node.

b) Secondary contradictions are those that arise between the nodes of a simple activity system. For instance, there may be contradictions between goals of the teacher and the influence of the school community on the final construct of the assessment tool. For example, the goal of the teacher may be to include test questions that require high cognitive demands and difficulty levels but the senior teachers and managers are reluctant to include questions that may negatively affect the pass rates of learners.

c) Tertiary contradictions describe potential problems emerging in the relationship between the existing forms of an activity system and its potential, more advanced goal and outcome. The advancement of an activity system as a whole may be undermined by the resistance to change, demonstrated by the existing organisation of the activity system. There could also be gaps in the knowledge levels of organisations or individuals to achieve the desired outcome. For example, a goal of the national ministry of education may be that teachers in their SBA design test questions that require high cognitive demands but the teachers are unable to achieve this because they have not been trained by the ministry, or have the knowledge resources, to do this.

d) Finally, quaternary contradictions refer to contradictions within a network of activity systems, that is, between an activity system and other activity systems involved in the production of a joint outcome. For example there could be contradictions between the SNA and SBA activity systems in relation to the assessment of mathematics. The SNA tools could be designed on the expected curriculum competencies whereas the SBA tools could be based on historical learner competencies.

Contradictions can therefore exist at various levels of the activity system – within each node of an activity system (for example, tensions within the subject), between nodes (for example, between the subject and rules) or between different activity systems (for example, between the workplace and university) (Barab et al., 2004). Recent research studies employing AT concepts identified contradictions and tensions occurring even in similar contexts. The identification of contradictions in an

activity system helps practitioners and administrators to focus their efforts on the root causes of problems (Engeström, 2001). Contradictions need not necessarily be viewed in a negative light but as problems requiring solutions (Issroff & Scanlon, 2002).

3.5 Conclusion: Applying an AT framework to this study

In this chapter, the discussion was structured to provide an introduction to, and information about, AT so that its application in this study could be understood. As indicated earlier, the preference in using AT over other socio-cultural theories was the rich attention given to the relation between individual and collective ideas, and the relationship between context and activity.

Of significant importance to this study was the understanding and application of AT concepts to provide an understanding on the structure of rules and how artefacts operationalise advocated rules within and across SNA and SBA leading to expanded explanations on contradiction identification, hierarchical activities and interacting activity systems. In particular, AT assists in identifying tensions and contradictions and provides a means of explaining the ways in which they play out. Barab et al (2002) identified the following three principles in AT that can be used as a frame for an analysis of an activity system: 1) structure levels of activity hierarchically; 2) characterise sub-components; and 3) locate points of contradiction. Each of these is briefly summarised.

3.5.1 Hierarchical structure of activity– motives, goals and operations

In this study, advocated rules were considered in the context of motives, goals and operations. Advocated rules were seen to be operational design features influenced or directed by broader motives and goals. SNA motives and goals led for the creation of “rules” (design features) for both SNA and SBA. The tension in this hierarchical structure is that broader motives and goals in these systems do not always align and that “tools” do not always follow “rules”.

In this study, the enacted rules were evidenced in the assessment artefacts. All the SNA and SBA documents (tests, frameworks, and associated documentation) were

viewed as assessment artefacts within their respective activity systems. The analysis of them focused on establishing the nature of the rules that these artefacts were oriented towards. It has been mentioned that some of these artefacts overlapped between SNA and SBA activity systems and the contrast of rules identified in the SBA context could be extracted in relation to an analysis of national policy documentation and the ANA assessments artefacts.

3.5.2 Characterise sub-components for SNA and SBA

By attending to the primary components of Engeström's (1987) EMT: Subject(s); Tools, Object(s); Outcome(s) Rules; Community; and Division of Labour; an investigator can begin to structure his/her analysis without the burden of too overt a prescription (Barab *et al*, 2002). Barab *et al*. (2002) found that often multiple analytic models are needed. In this study, expanding on Engeström's third generation model of AT, two culturally constituted 'collective practices' of assessment - the one being the internal assessment activities mathematics teachers generate and engage in their school-based programmes (SBA), and the other being the policy advocated standardised national assessments (SNA) are represented and described, firstly as individual activity systems and later in terms of their interacting activity systems.

Since the study was located in three different schools, the assessment activity system at each school needed to be looked at both as an individual system and as interacting sub-systems within a broader mathematics assessment system. Engeström's models of interacting activity systems provided a schema to analyse and compare components of the two activity systems within the context of a shared objective (see **Figure 3.6**).

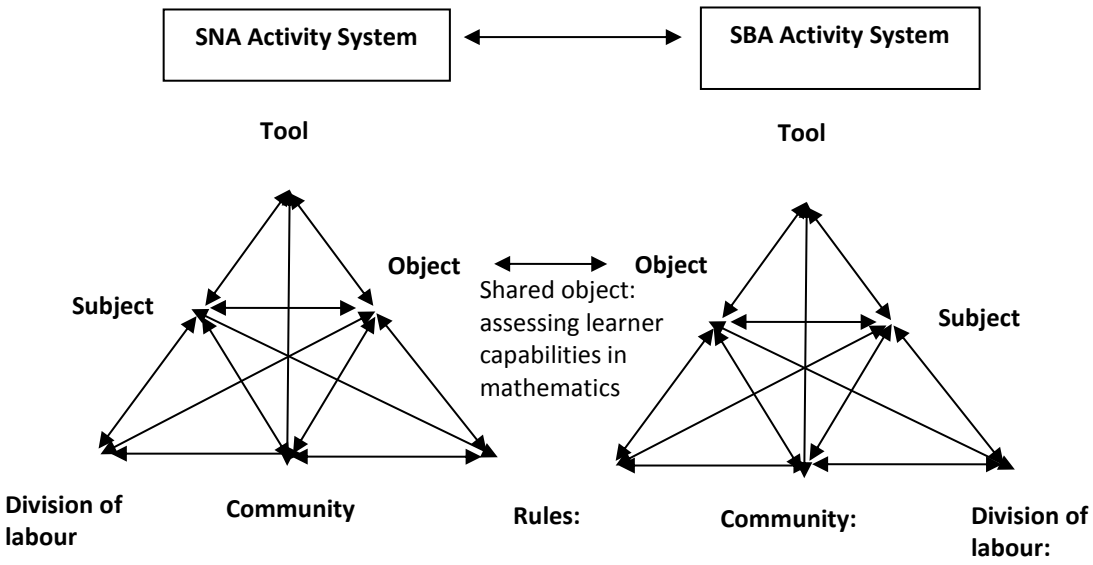


Figure 3.6: Interacting assessment activity systems

Figure 3.6 is an illustration of the model of two interacting assessment activity systems. Elements of the model were matched to corresponding features of Engeström's minimal models of interacting activity systems indicated earlier. Using these models an analysis of teachers' assessment activities was undertaken to understand a) how teachers individually design their assessments; b) the differences that exist between teachers' assessment practices and artefacts and how they compare with the policy driven SNA practices and ANA. While the study centered on the upper triangle of relations between the tools, subject and object, it was understood from an AT perspective that these relations are not independently constituted but are mediated by various factors including community, rules and division of labour (Barab *et al*, 2002), described by Jaworski and Petoni (2009) as the hidden curriculum ideas.

The EMT was used to define the context of assessment activity in each school. This involved an examination of the relations between Grade 6 mathematics teachers from three different mathematics centres (schools) and their SBA of learner competencies in mathematics mediated by the primary components that constituted the assessment activity system from the perspective of: (a) their assessments over the course of a year, and associated school level policy and supporting documentation tools (both created and mediated from other levels of the system), (b)

the overall assessment profile of teachers, (c) division of labor (assessment roles), and (d) rules (informal, formal, and technical).

3.5.3 Identify the level and nature of contradictions

The idea of *contradiction identification* was an important methodological issue for this study. It implied looking beyond a purely descriptive comparison of different assessment activity systems that featured in the assessment landscape of teachers to an expanded view of activities that were in conflict with each other. The nature of the conflict could be identified as primary, secondary, tertiary or quaternary contradictions between systems. It allowed for an account of how the teachers' activity systems located themselves across different hierarchical contexts: classroom, school, district and national. The global context was considered only in reference to these four. It also allowed for a richer discussion on the structure of the components both at the policy level (advocated rules) and at the teacher level in designing mathematical assessments. Finally, it also allowed for tertiary and quaternary contradictions to be identified and discussed when a new external object such as the ANA was introduced into the assessment activity system of mathematics teachers.

In the next chapter on the research methodology, attention is given to explaining how AT concepts are used to address the research questions of the study and how these research questions link to data sources.

4. The Research Design

In this chapter I explain the research methodology followed in the study. The discussion details processes followed in the research design including, sampling, data collection, document analysis and the interview process. Data drawn from the document analysis and interviews held with teachers revealed that both data sources were required to make substantive judgements against the research questions of the study. The analytical frame and how it is used to categorise data is also explained in this chapter. This chapter was written to foreground the structure of the analysis and findings that follow in the subsequent analysis chapters on SNA and SBA and in the conclusion.

In this study an analysis is presented of SNA and SBA as activity systems. Within each system, there were two broad analytical levels to consider. Firstly, the SNA and SBA policy documentation were analysed to understand the commonly advocated design rules for SNA and SBA and situate them in terms of Morgan's assessment discourse framework. Expanding on the discourse analysis, an EMT analysis then looked at specific components of SNA and SBA activity systems to better understand the relationships between advocated rules and their underlying motives and goals, and between advocated rules and tools. Secondly, the analysis looks at the specific test items in both systems to establish the alignment between advocated rules and enacted rules as seen in assessment artefacts (tools). In this part, a range of themes, categories and sub-categories were developed to conduct a detailed analysis of artefacts.

The comparative methodology applied in the study was to look at advocated and enacted design features within SNA and SBA separately and identify the disjunctures within and between them. The activity systems of SNA and SBA were viewed as two complimentary and interacting activity systems. Ethical considerations for this type of study are presented at the end of the chapter.

4.1 The research design

In general, this research study can best be described as naturalistic inquiry, with theoretical and grounded interpretations based on qualitative data (Guba & Lincoln,

1983; Scriven, 1983; Stake, 1983) generated through document analysis and interviews held with participants. As this study was exploratory and interpretive, data collection and analysis were to a large extent determined by the material evidence contained in teachers' assessment portfolios, the perspectives of the selected teachers in interviews and my understanding of the policy context on assessment.

Researchers working within the qualitative research paradigm investigate the quality of the relationships (e.g. between teachers) and materials and try to understand situations (e.g. assessment practices) (Fraenkel & Wallen, 1990). Qualitative researchers usually work in an exploratory-descriptive way and use sensitising and empathising rather than quantifiable concepts (Booyse, 1993). Merriam (2002, p. 4-5) suggests four key characteristics of qualitative research: a focus on understanding how people interact and experience the world and the meanings they have constructed, the use of the researcher as the primary tool for data collection, the use of induction to build concepts, hypotheses, and theories, and the production of complex and layered descriptive data. This is not to say that this study operated exclusively without any consideration to the use of quantitative or mixed methods to illuminate points of interest that arose in the study.

The overall research approach for this study was a case study design. Case studies comprise important and useful means of studying educational settings in qualitative research (Booyse, 1993). The interpretive nature of case studies allows the researcher to 'study and give insight into specific situations or events' (Stake, 1995). The defining aspect of a case study is the delimiting of the object of the study (Merriam, 1998). In this study, the phenomena being studied are the test design practices of mathematics teachers operating within a national curriculum and school-based assessment context in three different settings. Merriam (1998) argues that 'cases' are studied in their own right, not just as samples from populations. In this way, the generalisation of findings to a broader population or community is not a primary issue of concern in case study designs. Fraenkel and Wallen (1990, p.370) state that much can be learned from 'studying just one individual, one classroom, one school or one district.'

A case study design was an appropriate research approach in this study as it

allowed for an exploration of the mathematical assessment activities employed by the teacher in each school, discovery of observed interactions, identification of contradictions and variations in relation to assessment practices, and understanding of the different contexts within the assessment system of the teachers. Bassey (1999) states that the 'great strength' of a case study design, is that it allows the researcher to concentrate on a specific instance or situation and to identify interactive processes at work (which may not be evident in large scale surveys). I incorporated a comparative focus across the cases in this study in order to follow up the consistent finding of differential performance of schools drawn from different socio-economic backgrounds in South Africa.

The value of this case study was in its relatability to other teachers assessing mathematics in the context of interacting assessment activity systems. Relatability in the context of this study would refer to other Grade 6 Mathematics teachers who are participants in this dual assessment activity context of SNA and SBA. Bassey (1999) argues that the relatability of a case study is of more value than its generalisabilty.

4.2 Sampling

The sampling design was a purposive selection of three co-educational public schools in the three different educational setting types prevalent in the Gauteng province: a township, an inner-city and suburban area. School type was a key consideration within the selection of the three schools. By looking at schools in different settings, more detail was provided on the nature of and the extent to which different school communities, policy layering and mediation, and teacher role contributed towards the teachers' design and use of assessment tasks. The settings were three schools were located within three different educational districts in the province. Historically the township school had fewer resources than the suburban and inner-city schools and had been regarded by the local district as an underperforming school.

The principal participants in this case study were Grade 6 Mathematics teachers from the three selected schools. Grade 6 teachers that were familiar with the assessment requirements of the RNCS, the FFL milestones and the ANA were asked to volunteer as participants in the study. As indicated in Chapter 1, teachers'

assessments were looked at both as individual constructs and as artefacts within a collective Grade 6 Mathematics community constrained by access to knowledge and resources. Opportunities existed in the study to examine commonalities and variations in the design of SBA in schools from a township, an inner-city and suburban area.

Selection of these teachers was also based on the interactions I had with them in workshop training and interviews held with principals during on-site visits to the schools within my role as a systemic evaluation monitor from the DBE. The willingness of principals to allow teachers to participate in the research was an important determinant of school and teacher selection. The schools selected had functional leadership structures and available artefacts that indicated that teaching and assessment activities occurred regularly, which were important in the context of identified dysfunction in these aspects (Taylor & Vinjevold, 1999). The feasibility (distance and cost of transport) for the duration of the study was also taken into account. Finally and perhaps most importantly, the willingness and cooperation of the teachers for the entire duration of the data collection (12 months) was required for the case study to be of value.

My first interaction with the teachers was done telephonically in July 2009 to set up an informal meeting. These meetings took place in the first week in August at the schools of the selected teachers. At this meeting, teachers were briefed on the focus of the research and why I considered them to be appropriate participants for the study. They were briefed on the types of data collection techniques to be used, the duration of study, and the commitment and willingness needed for the study to be meaningful. At the respective meetings, the three teachers (whom I henceforth refer to as Kalay – from the suburban school, Fiona from the inner-city and Mary from the township school) agreed to be volunteers and were generous and eager to share information on how they planned, constructed and delivered their mathematics assessment tasks in the classroom. The teachers had no objections to the data collection techniques as long as I obtained permission from the senior management of the school to analyse teacher records, learner workbooks, and conduct interviews with them. In Chapter 6 more information about the background, experience and practices of the teachers concerned is provided.

Subsequent to the meeting held with the teachers, a meeting was arranged with the principals of the schools and this took place a week after I had met with the teachers. I had prepared a formal letter (see **Appendix 1**) requesting permission to conduct research at the school with the selected participants. In the letter, I also explained the nature and expected duration of the research and how at the end, the study could benefit the teachers and the school. At the meeting with the principal, permission was granted. In August 2009, I also requested permission from the Provincial Department of Education (Gauteng) to conduct research at the three schools and informed the relevant units on the nature and duration of the study. Data collection was planned with the consent of the participants for September and December 2009 for the first interview followed by a second interview six months later during June to July 2010. Data on the teachers' assessment tools and design emanating from the document analysis conducted during 2009-2010 was triangulated with data collected through the two interviews.

4.3 Data collection

In qualitative research triangulation of data is used as a technique of pattern seeking (Schumacher & McMillan, 1993). Denzin (1978) cited in Schumacher and McMillan (1993) refers to triangulation as the cross-validation among data sources, data collection strategies, time periods and theoretical schemes. Through triangulation, the researcher is able to find regularities in the data and the reliability of findings are improved. An important consideration for me was that information collected should be triangulated across the two intended techniques to be used: interviews and document analysis (see **Figure 4.1**). In terms of my study, interviewing allowed me to get teachers' subjective views for their selections and sequencing of internal assessment tasks, as well as insights into the goals, school rules, communities and divisions of labour within which their internal assessment activities were located. Document analysis of SNA and SBA tools provided the initial evidence from which points of interest were further explored during the interview sessions.

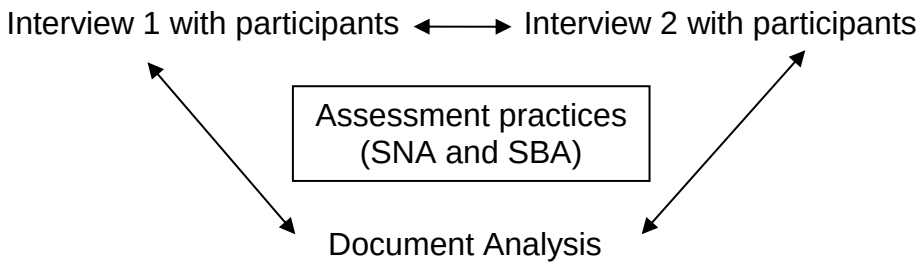


Figure 4.1: Triangulation of data

(Adapted from Schumacher and McMillan, 1993)

An important process in the development of the data recording tools was their refinements after they were piloted with other grade 6 teachers (not from the same schools selected in the sample). The piloting of the interview schedules took place in two schools that did not form part of the main study. The pilot process was necessary in the narrowing of categories chosen in the above tools and for a sharper focus on design features. The document analysis was narrowed to specific records of the teacher which would render appropriate data for me to establish patterns against the research questions. Recording and reporting tools, pass rates, feedback to parents and completion of support forms for underperforming learners were treated as background knowledge. Similarly, the SNA focus was limited to the design feature rules and not so much the administrative conduct and logistics of large scale assessment studies.

In piloting the interview instruments, new information came to the fore in terms of the role of the district in mediating SNA artifacts and the varied in-school support teachers received on moderation, peer assistance, and access to resources. Categories were then added to the final instruments to explore these aspects in greater detail. Technically, the piloting of the interview schedules helps reduce inconsistencies in responses and sharpens the focus of questions on issues that need clarification (Williamson, 1982).

The data collection instruments were interlinked (as they were informed by the same research questions) and both informed emerging patterns, results and conclusions. The triangulation of the data becomes more evident in the analysis and findings that are the focus of the next two chapters. The data collection instruments were linked to specific data sources in the document analysis and interviews. The discussion that

follows explains the methods followed in generating evidence for the document analysis and interviews, linking the data sources to the analysis, and describing the categories used in the coding of the generated data. The document analysis is explained first followed by a discussion of the interview process.

4.4. Document analysis

The first phase of data collection and analysis involved a comprehensive document analysis of the SNA and SBA artefacts relating to the assessment practices of the Grade 6 Mathematics teachers involved in the study. Document analysis, or content-analysis, as its name implies can be used to analyse past and present records of the participants themselves and/or the system under investigation. Fraenkel and Wallen (2011, 1990) argue that a "person's or group's conscious and unconscious beliefs, attitudes, values and ideas are often revealed in the documents they produce." In this study information gathered from policy documents and interviews was triangulated with information obtained from official and unofficial records of the teacher. Van Daler (1973) has argued that personal documents, such as diaries and letters and various other physical, psychological and sociological reports (as they appear in teacher records) yield valuable information.

The document analysis was structured into two parts to coincide with the overall research focus, namely, 1) the design rules and 2) the assessment artefacts. The document analysis took into consideration both the broad literature's focus on these two aspects and the empirical evidence arising out of the local contexts of SNA and SBA.

4.4.1 The design rules

As indicated earlier, the analytic approach for the advocated design rules was to first look at a hierarchical structure of rules influenced by motives and goals and then to look at the nature of rules through an EMT analysis of SNA and SBA activity systems. In SNA, the hierarchical structure explicated the policy context driving assessment in South Africa while for SBA, it provided a sense of how teachers perceived factors influencing their SBA activities. To better understand the underlying influence of motives and goals on rules, I made use of Morgan's

assessment discourse framework to classify motives and goals in terms of her discourses: psychological, curriculum implementation or curriculum standards.

The EMT analysis allowed for an understanding of the structure of each school system, the relationships and tensions within each one, the variation across schools and how these differed from SNA. As indicated in Chapter 2, assessment tools exist within specific policy contexts with associated goals, role players and rules. By looking at specific elements within the activity system where the assessment tool is located, valuable insights were gained on the nature of actual assessment tools against what was advocated.

The data sources for understanding advocated rules in the SNA context were the policy artefacts influencing assessment with a particular focus on the ANA as the key assessment artefact. Policy artefacts (as they are explained in Chapter 5) included the RNCS, FFL milestones and education sector plan. Broader SNA artefacts linked to the ANA policy context included published reports, guideline documents, frameworks, and the ANA Grade 6 Mathematics assessment tasks administered in schools during 2008 to 2010. Analysing and describing the policy context of SNA provided an understanding of the nature of the policy imperatives driven by the state and its impact on the design of the ANA as a key artefact.

In the SBA context, the advocated design rules were sourced from the policy prescripts on assessment and planning documents of the teacher. Advocated rules for each teacher across the three selected schools were documented along three hierarchical levels of influence: 1) national policy texts, 2) district texts, and 3) school (including teacher) texts. The intention was to establish evidence of national, district and school policy directives on teachers' assessment tasks.

In SBA, the key policy documents analysed at the national level were the RNCS and supporting guidelines and FFL assessment frameworks. District policy prescripts included documentation that was received by the school and teachers through meetings, workshops and circulars. School level documentation was analysed for what teachers included in the assessment portfolio files. This provided sources of motives and goals underpinning the school rules as these documents provided the

context for the continuous assessment planning cycle of the teachers.

The district level formed an intermediate level within which national artefacts were mediated for school based teachers. While the main focus was on understanding the interaction between the SNA and SBA systems, the analysis also took into account the influence of district and school mediated policy prescripts located within the broad policy mandates intended to increase the understanding and practicability of national assessment policies for classroom implementation. The analysis of materials from district workshop and cluster meetings involving Grade 6 Mathematics teachers gave valuable insights into the role played by the district in relation to interventions, support programmes and external moderation procedures. In AT terms, the role played by the district expanded the community of the teacher beyond the school and in some instances (e.g. the township school) had a great impact on the division of labour.

At the school level, documentation was also categorised according to whether the source was national policy texts (such as the RNCS and FFL documents), district texts (including circulars, notices, moderation forms) and school texts (evidenced by school level documentation and textbooks). Particular attention was given to the role of textbooks in the design of assessment tasks by the teachers. This provided information on the resources that the teachers considered valuable for the purposes of assessment in their SBA. It was also important to understand how teachers assessed within the rules applied in their school community. Assessment practices in this category were described as school level support and moderation. Within the three schools, the analysis of rules included broader rules related to the school's moderation procedures, the quality assurance of tests and feedback forms given to the teachers.

4.4.2 The assessment artefacts

The ANA and SBA tasks and tests formed the key assessment "tools" to examine, The approach to examining these assessment artefacts began with a focus on 3 key themes: 1) mathematical coverage, which was related to the design features identified in the literature on mathematical domains related to the curriculum; 2)

range and scope of questions, which was considered in relation to common design features identified (namely, difficulty level, cognitive demand, language demand, and item format); and 3) a third theme, the assessment of common topics, was added into the SBA analysis to gain more in-depth insight on content selections. The third theme was necessary because the categories drawn from the literature, in the SNA context, were too broad to compare schools in terms of the breadth and depth of selections from specific mathematical topics within assessment tasks.

Actual school test artefacts included formal examination tests, class tests and activity-based assessments. Tools designed by the teacher were matched to the different types of assessments and their rules. Within the activity based assessments, further analysis was done on mathematics topics from a sample of assessments that featured commonly across the assessments designed by all three teachers. This provided a way of highlighting similarities and contrasts between the three schools.

Design features and sub-categories

The themes, design features and analysis are summarised in **Table 4.1**.

Table 4.1: Themes, design features and sub-categories

Theme	Design feature	Sub-categories				
Coverage	Mathematical domain (MD)	LO 1: Number	LO 2: Patterns	LO 3: Shape and Space	LO 4: Measurement	LO 5 Data
Range and scope of questions	Difficulty level (DL)	Easy		Moderate		Difficult
	Cognitive demand (CD)	Knowledge	Routine	Complex	Non-routine	
	Language demand (LD)	No text		Low text		High text
	Item Format (IF)	Closed short answer (CSA)		Multiple choice questions (MCQ)		

These common design features allowed comparisons to be made between the assessment artefacts within SNA and SBA contexts. To investigate the nature of these design features, it was necessary to develop sub-categories with clear descriptors, once again drawn from the literature. From the sub-categories, analysis tables could be generated. The analysis tables were then located under specific

themes to provide a basis for explanations and for comparisons to be made across artefacts in SNA and across the three schools in SBA, and later to compare the analysis tables generated in SNA with those in SBA.

Theme a: Coverage

Analysis of coverage, in different ways, has a long history in studies of Opportunity-to-learn (OTL) in relation to curriculum incoherence (Reeves & Muller; 2005) and in studies on alignment of assessment, standards and instruction (Porter, 2002). In **Table 4.2** an indication is given of the design feature and elements for mathematical coverage. Coverage of LOs and ASs emerged as relevant aspects from the national curriculum to provide a detailed explanation on coverage issues in both SNA and SBA.

Table 4.2: Tool A: Mathematical coverage in an assessment task

Design Features	**Element 1**	**Element 2**
Mathematical domain (MD)	Coverage of LO 1–5:	Coverage of individual ASs per LO*

* In Chapter 2 the expected individual AS count per LO was listed.

Since I had chosen to look at a full academic year of assessment for each teacher, a necessary sub-category was the extent to which the different mathematical domains listed in the RNCS as Learning Outcomes (LOs) were covered. In Chapter 2, these were listed as: number, patterns, shape and space, measurement and data handling (DoE, 2002b). The mathematical domain category was also used to establish the extent and nature of the coverage of ASs within a LO that featured in the ANA and in teachers' assessment tasks. This was useful for making comparisons on curriculum coverage between the SA-SNA context and the SBA of teachers.

The focus on coverage also led me to look at frequency counts of assessment standards (ASs) from individual test items. I counted individual mathematical skills (listed as the ASs) across the assessment tasks and established their frequency of occurrence across the ANA tests in SNA and across the SBA tasks of the selected teachers. The frequency was based on non-repetitive counts of ASs within individual assessment tasks. Non-repetitive counts were considered because I was more interested in the spread of individual ASs covered rather than the number of times an

individual AS was repeated in a task. In the SNA context, frequency counts were done on the ANA across 2008–2010 to build a picture of the nature of mathematical coverage in the ANA over time. In the SBA context, the count was limited to the 2010 academic year assessments. Counts of the actual number of test items in each the ANA tests and in the end-of-year formal examinations of teachers were also recorded.

Theme b: Range and scope of questions

The range and scope of questions were based on the design features related to Difficulty Level (DL), Cognitive demand CD, Language demand (LD) and Item Format (IF). For each design feature, descriptors were developed based on the literature reviewed and their historical use in the ANA. A literature review of these features was presented in Chapter 2. The same descriptors were applied to the assessment tasks of teachers in the SBA context.

By considering the DL of individual questions posed in assessment tasks the overall spread of difficulty could be established. Since the DL analysis was done according curriculum competencies relevant to a specific grade, it provided valuable information in relation to South African literature arguing that teachers often design assessment tasks for minimal levels of achievement by learners (van der Berg, 2006). The CD of test items revealed the extent to which different levels of mathematical thinking featured in internal school tests. It has been argued that tasks should operationalise the full range of mathematical thinking goals. Anderson's (2005) revision of Bloom's (1956) Taxonomy of Educational Objectives provided a two dimensional Taxonomy Table to classify educational objectives into specific categories that describes cognitive processes. The CD categories in this study are congruent with Anderson's taxonomy as well as those used by the DoE (2009b) in their examination guidelines to schools and in their curriculum and assessment policy statement (CAPS) (DBE, 2011c).

By looking at the LD of questions teachers selected, I was able to look at the level of language teachers included in contextual applications and in the instructions used to assess mathematical skills. This was significant as language has often been seen by research commentators as a key inhibiting factor in South African learner

performance in national assessments (Howie, 2002). Finally the IF of test questions was classified in terms of multiple choice questions (MCQ), and closed short answer (CSA) questions to establish the extent of offering learners varied and multiple opportunities to present their mathematical skills. This choice of design features was aligned to and extended those used in the ANA and the international assessment studies across SNA and SBA.

Theme c: Assessment of common topics

The third theme added in the SBA analysis was aimed at understanding the decisions teachers made in their design of assessment tasks and selection of mathematical content in relation to common topics. This theme allowed for an in-depth comparative qualitative analysis of the range and scope of questions across the teachers. As I engaged with the analysis, it became important to also consider certain factors (e.g. use of textbooks) that seemed initially to be more peripheral to mathematical content selection, but later came to be understood as essential for a more complete picture of the selection process, and thereby enhancing the understanding of the mathematical coverage and the range and scope categories.

I referred to this theme as the assessment of common topics of teachers as it gave insight into the preferences of each teacher on content selections within specific topics. The categories within this theme initially emerged out of the comparative document analysis and were explored in greater detail in the interview data. This theme provided valuable insights on the relationships between school context and teacher choices when designing assessments on specific topics. Common topics were identified by looking across the assessment portfolios of the three teachers. Although I anticipated drawing comparisons on all common topics across all five learning outcomes listed in the teachers' portfolios, differences in the nature of work covered by the selected teachers did not allow for comparisons to be drawn within each of the five LOs. The common topics that were selected were within LO 1, LO 4, and LO 5. These are indicated in **Table 4.3**.

Table 4.3: Selection of common topics

LO	**Topic**
1 – Number	Ratio and Rate.

4 – Measurement	Capacity.
5 – Data handling	Graphs: Bar graphs

These topics appeared in the portfolio files of all three teachers, linked to the same identification of LOs indicated by the teachers. Further, assessment activities on these topics were consistent with evidence in learners' work books. A narrower yet sharper focus on common topics was preferred to understand the assessment activity systems of teachers in greater depth. After completing the document analysis, matters that required further explanation and understanding were identified and explored through the interviews held with the teachers.

4.4.2.3 Item maps

The mathematical coverage and range and scope themes and sub-features allowed for summary coding of assessment item characteristics in SNA and SBA (see **Table 4.4**). Descriptors for each of the sub-categories are presented and discussed in Chapter 5.

Table 4.4: Item characteristics for the selection of mathematics

Categories	Descriptor	Descriptor Code
Mathematical domain: (LO and AS)	Number, operations and relationships	LO 1 (AS 1.1 – 1.12
	Patterns, functions and algebra	LO 2 (AS 2.1 – 2.6)
	Shape and Space (Geometry)	LO 3 (AS 3.1 – 3.8)
	Measurement	LO 4 (AS 4.1 – 4.11)
	Data Handling	LO 5 (AS 5.1 – 5.10)
Difficulty level (DL)	Easy	E
	Moderate	M
	Difficult	D
Cognitive Demand (CD)	Knowing basic facts	K
	Applying routine procedures	R
	Using complex procedures	C
	Solving non-routine problems	N
Language Demand (LD)	High text	HT
	Low text	LT
	No instructional text	NT
Item Format (IF)	Multiple choice questions	MCQ
	Closed short answer response questions	CSA

These item characteristic codes led to specific test item maps. For each test or

assessment selected, the process required each item or question to be allocated item characteristics based on the categories listed. For example, an exemplar item map listed for the first 10 questions of the 2008 ANA in the SNA category is listed in **Table 4.5**.

Table 4.5: Exemplar item map for SNA

2008 ANA		Question Number										
		1	2	3	4	5	6	7	8.1	8.2	9	10
Category	LO	1	1	1	2	1	1	2	1	1	1	1
	AS	1.8	1.10	1.10	2.5	1.8	1.8	2.1	1.4	1.4	1.3	1.4
	DL	E	E	E	M	M	M	M	E	E	M	D
	CD	K	K	K	R	R	R	R	K	K	R	C
	LD	LT	LT	LT	LT	LT	LT	LT	LT	LT	LT	HT
	IF	CSA	CSA	CSA	CSA	CSA	CSA	CSA	CSA	CSA	CSA	CSA

For question number 1, the item characteristics can be read as follows:

Question 1	LO 1 – Number
	AS 1.8 – Estimates and calculates by selecting and using operations appropriate to solving problems that involve rounding off to the nearest 10.
	DL – Easy
	CD – Knowledge
	LD – Low Text
	IF – Closed Short Answer

The data sets for analysing the mathematical coverage and range and scope in SNA and SBA were structured on these comparable item maps. In the SNA context, item maps were generated for each assessment year for the 2008, 2009 and 2010 ANA tests. In the SBA context, item maps were generated on the SBA for Kalay, Fiona and Mary. In Chapter 7, graphical summaries from the item maps are used to compare the selections of mathematics in SNA with SBA

4.5 The interview process

As indicated earlier, two rounds of interviews were held with teachers over a six month period. The purpose of the interviews was to understand the teachers' justifications for designing and selecting assessment tasks and to triangulate evidence from documented texts on assessments they had designed.

In AT, specific values are accorded to interview data. While, according to Nardi (2007), it has become a kind of received wisdom in the research community that people cannot articulate what they are doing (a notion sometimes used as a justification for observational studies and sometimes used to avoid talking to users at all), there are also indications that carefully structured interviews make it possible for the researcher to overcome some of these difficulties. In this study, whilst I could construct an assessment activity system "from above" based on documentary evidence collected, I wanted to gain insights into subjective views about assessment activity within the AT systems of SNA and SBA. Given my professional role in the national department, understanding conditions and imperatives in these subjective ways was important for understanding contradictions within the SNA policy context itself and between SNA and the SBA practice of teachers.

Data collected through the document analysis of teachers' files and pre-planning was matched and checked against data collected through the interview sessions held with participants. In addition, participants were called upon at various times to validate results, discussion and interactions as the study progressed. Since interviews allowed me to explore, inquire and interpret issues as they arose in the study, it became the principal subsequent mode of data collection, both informing and informed by patterns and categories emerging in the document analysis. Using a structured interview methodology (Posner & Gertzog, 1982; Schumacher & McMillan, 1993), two interview schedules were developed. The first interview schedule is included in **Appendix 3**. The interview process included the following stages: early planning, drawing up the questions, piloting the interview schedule, conducting the interview and analysing the data.

The interviews took place separately in each of the schools at a pre-arranged time. The first interview schedule was designed to ensure that the same questions were asked to all respondents to ensure comparability (Fraenkel & Wallen, 1990). In the first interview, the focus was on understanding the level and nature of interaction across national/district/school texts. The emphasis here was to elicit data on:

1. How teachers engaged with policy texts (e.g. FFL milestones) and how these texts influenced their design of assessment tasks

2. The use of standardised tests (e.g. district common papers, ANA)
3. How resources (e.g. textbooks) play a role in designing assessments.
4. What professional support was available at the school?
5. What guides the selection of content in the assessments?
6. Individual planning for assessment
7. Design preferences on their assessments

In the second interview schedule, the broad outline was the same but questions were more task-specific to the individual teacher. Participants were probed through specific questions on their individual assessment tasks. In both interviews, a mixture of open-ended questions and direct questions was drawn up to maximize respondent participation (Thompson, 1978). In the second interview, the purpose of the interviews was to understand the range of skills assessed and the nature of mathematics covered by individual teachers. The emphasis here was to elicit data for the selection of mathematics analysis. The tasks selected were based on a selection from a range of test categories that included: summative moderated tests/examinations (set either internally or externally), summative class or cycle tests, and formative/class based activity tasks. As a means of comparing teacher assessments from different school settings, assessments done by the teachers on a sample of common topics were also included and probed in the interview. In each type of assessment, the teacher was asked to provide a sense of:

1. How each test was developed/designed and for what purpose?
2. The standard of the test.
3. The categories (or classification criteria) used in the design of the test.
4. The reasons for the choice of the number of questions and sub-questions on a certain topic.
5. The range and build-up of sub-questions.
6. The intended difficulty of a question. Teachers were asked to indicate which questions they considered more or less difficult, and why
7. The language and contexts used.

8. The balance of different cognitive type of questions (knowledge recall/ non routine.
9. The allocation of marks.
10. The main influences underlying their generating of the test.

In both interview schedules, questions fell into two broad categories: 1) general questions on assessment and 2) specific questions relating to the school based assessments of the individual teachers. The responses were elicited through a 'face to face' conversational interaction. The data generated was of a nominal type. According to Cohen and Manion (1991) the advantage of this type of response mode is that it allows for less biasing and greater flexibility but the disadvantage is that it is more difficult to code.

In conducting the interviews the purpose of the research was explained (Williamson, 1982) and respondents were allowed to ask clarification questions during the interview (Osborne & Gilbert, 1980). Respondents were assured of their anonymity throughout the interview process. In analysing the two rounds of interview data in this study, the responses were interpreted in the light of the research questions indicated in Chapter 1. Questions were categorised and clustered into subcategories that were verified with professional colleagues and my supervisor (Cohen & Manion, 1991). Responses from the selected teachers were not accepted at face value but checked and backed up with evidence from other data sources (triangulation) for consistency and validity (Cohen a& Manion, 1991; Schumacher & McMillan, 1993; Williamson, 1982). For example, assessment tools evident in the teachers' portfolio files were checked against actual assessments learners included in their workbooks.

Detailed interview transcripts were generated for each round of interviews conducted with Kalay, Fiona and Mary. Coding according to Cohen and Manion (1991) has been defined as the translation of question responses into specific categories for data analysis. In this study, summaries of tape-recorded interview transcripts were later coded according to pre-identified categories established through the document analysis. An exemplar of coding a teacher's formal examination task is included in **Appendix 4**. Responses were then grouped and rank coded to determine how often

particular patterns recurred. Finally the data was analysed and interpreted in light of the research objectives (Cohen & Manion, 1991). Throughout the interview stages, processes were not “cast in stone” and were flexible and interchangeable. Continuously checking and reflecting on what was planned provided me with valuable insights on the relevance and context of the emerging data in relation to the study’s research questions.

4.7 Ethical considerations

A final aspect of the research process was to understand the ethical considerations of engaging in such a research study. Key among these considerations was that the findings should be reliable and valid based on the data generated through the research instruments. Lincoln and Guba (1985) state that a qualitative researcher must strive to ensure that the findings and interpretations of the research (the researcher‘s constructions) are actually credible representations of the participants’ understandings and experiences (participant constructions).

Several measures were taken in the research process to ensure and improve validity of findings. Firstly, at the beginning of the study, piloting the research instruments provided a check of whether the data collection tools were appropriate and what kind of data would be generated through them. Secondly exploratory meetings were held with principals of selected schools to identify whether the school environment was conducive for research activities to be carried out and sustained during the entire research period. In this regard, the teachers were willing participants and consent was obtained from both the principal and the selected teachers (see **Appendix 1 and 2**). Thirdly, the teachers were afforded opportunities to confirm the accuracy of information provided through follow up interviews and telephone conversations, and further triangulation between data sources provided descriptive and interpretive validity (Maxwell, 1992).

Researchers point out that in doing a qualitative research, there is a ‘human factor’ which includes some unknowns so hundred percent external validity is unachievable (Scaife, 2004). Maxwell (1992) argues that understanding the research problem is more fundamental than validity for qualitative research (Maxwell, 1992). Thus, understanding the teachers’ assessment activity systems construed as an agreed

‘objective reality’ between me as researcher and the participants was a priority.

4.8 Conclusion

In summary, this chapter focussed on explaining the methodology of the research process followed in this study. The chapter detailed the critical processes followed in the research design, sampling, data collection, data analysis and ethical considerations. A critical part of this chapter was explaining the document analysis and the interview process foregrounding how data would be generated and analysed in subsequent chapters.

As explained in the document analysis, the methodology on design rules is based on a hierarchical structure of activity relative to motives, goals and operations and an EMT analysis. Operations are unpacked by looking at assessment artefacts in terms of design features, translated into item maps that facilitated comparisons of enacted design features across SNA and SBA, as well as comparison with advocated design features. The next chapter focusses on SNA in the South African context.

5. Standardised National Assessment in South Africa

5.1 Introduction

In this chapter a detailed analysis of SNA in the South African context is provided. I refer to this focus as the SA-SNA analysis and the discussion is dealt with in two parts. First, the "rules" for design advocated in policy texts are discussed, and second the SA-SNA artefacts are analysed in relation to what was advocated. The approach in the first part is to look at the hierarchical structure of SA-SNA where design rules are influenced by motives and goals and then to look into the nature of advocated rules by considering an EMT analysis of SA-SNA as an activity system.

As indicated in Chapter 3, activity can be looked at in terms of motives, goals and operations in terms of a hierarchical structure using the ideas of Leont'ev (1981). In the SA-SNA context, motives are looked at from the view of the State mandating collective activities for improving teaching and learning outcomes and can be found in national policy texts. Goals are seen as more specific assessment actions emanating from the collective motives and can be linked to Morgan's assessment discourse frame on psychological, curriculum implementation or curriculum standards. In the SA-SNA context, the ANA is considered as the key operational artefact and is viewed as having varied goals serving a broad educational community, while at the same time having a "backwash" effect into the rules of design in the SBA space. The EMT analysis extends the discussion of the nature of activity in SA-SNA and provides valuable information on design rules by exploring and examining tensions and overlaps between elements of the SA-SNA activity system and identifying points of contradiction. The documentary evidence for understanding "rules" in the SA-SNA activity system includes the RNCS document, the FFL assessment framework, the education sector plan, ANA assessment guidelines and frameworks, ANA reports released by the DBE, and the ANA tests.

In the second part of the chapter the analysis of SA-SNA focusses specifically on the assessment artefacts. The artefacts are analysed against the themes and design features of mathematical coverage and the range and scope of questions outlined in Chapter 4. To recap, these include mathematical domain coverage, the difficulty

level of test items, the cognitive demand of test items, language demand (amount of text included in items), and the item format.

5.2 The hierarchical structure of design rules in SA-SNA

5.2.1 Motives of SA-SNA

The discussion of motives focusses on three broader education policy imperatives that have shaped the design and purpose of national assessments at the primary school level in South Africa. These are the Revised National Curriculum Statement (RNCS), the Foundations for Learning (FFL) Campaign and the education sector plan referred to as the *Action Plan 2014: Towards the Realisation of Schooling 2025* (DBE, 2010). It should be noted that these policy imperatives have directives on assessment that relate to both SNA and SBA but the focus here is on their reference to SNA at the level of orientating goals to the SNA discussion.

The RNCS

According to a task team report commissioned by the Minister in 2009, the RNCS of 2002 did not revise the National Assessment Policy of 1998, maintaining the types of assessment and assessment principles of Curriculum 2005 (DBE, 2009). According to the RNCS assessment guidelines for mathematics (DBE, 2002b), there were five types of classroom assessments that could be utilised depending on the assessment purpose. These are summarised in the table below.

Table 5.1: The five different types of assessments listed in the RNCS

Type of assessment	Purpose and goals
Baseline assessment	Used to establish what learners already know at the start of a topic or mathematical section.
Formative assessment	Used to inform teachers about the progress learners are making within a topic or section.
Summative assessment	Used to give an overall picture of learners' progress at a given time, for example, at the end of a topic, section or school term.
Diagnostic assessment	Used to identify, scrutinize and classify learning difficulties so that remedial help can be provided.
Systemic assessment	Used to monitor externally the education system by comparing learners' performance to national indicators of learner achievement.

Source: Department of Education (2002): RNCS Assessment Guidelines for Mathematics

According to the guidelines, each assessment must be appropriate for the purpose for which it is used. The first four types of assessment relate mainly to SBA, although some forms of summative and diagnostic assessments could be designed and mediated by external agents (e.g. the local district office). The fifth type of assessment (systemic) relates specifically to SNA motives, where assessment tools are designed by the State to establish learner performance against specific national indicators. Depending on the nature of the national indicators, the design of systemic assessments may be closely aligned to the other four types of assessment. For example, a systemic assessment may have the same features as a summative end of year school examination if the goals of both are to assess learner competencies across the breadth of the curriculum. As suggested in Chapter 2, the ANA would fall within the systemic type of assessment, while at the same time embracing elements and goals of the other four.

Another key feature of the RNCS for designers of SNA was the outcomes based assessment (OBA) principles underlying the different types of assessment. These were common principles for all subjects relating to the outcomes based education (OBE) approach that focused on the holistic development of the learner. Some commentators argued that an OBE approach required South African officials and teachers to follow new "high-quality" principles of assessment that were more learner centered than pre-OBE approaches. However, many in the education sector were not well prepared for this paradigm shift that required changing their assessment practices (Vandeyar & Killen, 2003; Jansen 2001). **Table 5.2** outlines the DBE's (2007) OBA principles.

Table 5.2: Principles of OBA

OBA Principles	**Definition**
Design down	When planning assessment: • Identify the relevant learning outcomes and assessment standards. • Identify the skills, knowledge and values that need to be assessed. • Choose an appropriate assessment strategy. • Plan steps for differentiation to accommodate learners at different levels

Clarity of focus	learners should understand: • The criteria against which they are to be assessed. • The evidence of learning they are expected to demonstrate.
High expectations	Learners should be assisted and supported to reach their full potential. They should measure progress against their previous achievements and not against those of other learners.
Expanded opportunities	Learners should have multiple opportunities to demonstrate their full potential. We expect all learners to succeed but not necessarily at the same time and in the same way. Opportunities should be maximised for every learner by: • Taking into account different learning styles and multiple intelligences. • Presenting and enriching the curriculum in different ways.

(Source: DoE, 2003: Revised National Statement for Mathematics)

The description of the above principles affirms the literature commentary in Chapter 2 that SNA practices should satisfy the principles of reliability, validity, fairness, discrimination and meaningfulness (Vandeyar & Killen, 2003). However, there are contradictions and tensions that arise out of the above principles. It is noticeable that there is an expectation that the curriculum forms the starting point (as suggested by bullet 1 above) and teachers then adjust from there. This continues into the "Clarity of focus" principle, but the "High expectations" blurb is more learner-orientated, and works against the advocacy of beginning with the standardized curriculum that is seen in the first two principles. A recent curriculum initiative from the DBE in 2010 led to the introduction of a more streamlined standardised curriculum and assessment policy statement (CAPS). Because of its systemic nature, assessment tools such as the ANA also contradict the view of learner understandings being the starting point for assessment design.

The FFL campaign

The initial policy imperative for introducing standardised national assessments like the ANA was the Foundations for Learning Campaign (FFL) which specified intentions to support and regularly measure performance in literacy and numeracy. Linked to the FFL campaign, the ANA tests were designed to assess learner competencies based on the defined assessment standards of the RNCS.

The Foundations for Learning Campaign has been launched to focus the

system on improvement of learner performance in numeracy and literacy…The annual assessment of learners that will be set by the Department of Education will be based on these (FFL) quarterly assessment tasks. (DoE, 2008a)

State institutions (like the DBE) usually make their assessment features known through a listing of generic assessment principles that are meant to guide practice at classroom level while enforcing standardisation in a context of inequitable learning outcomes and accountability. With the introduction of the FFL campaign there were moves to a more detailed accountability framework for assessment at primary school level. Assessment milestones (knowledge and skills) were written to form termly units to assist teachers to develop the required assessment tasks per term. The structure of national assessment tools (e.g. ANA) was to be based on these quarterly assessments in mathematics and languages for grades one to six (DoE, 2008a).

The FFL policy, which made standardised testing of all learners compulsory for public schools, had precedents in the international context. In 2002, President Bush of the Unites States of America signed a school reform measure that required standardised testing of every pupil in mathematics and reading every year in Grades 3 through 8, so that assessment could be used as a tool for school improvement.

The objective of the FFL was to create a national focus to improve reading, writing and numeracy with set targets using the ANA as the key measuring tool. Provincial education departments (PEDs) responded to this call by drafting their own action plans for learner improvement and set targets for districts and schools. The Gauteng Department of Education (GDE), for example, in 2009 drafted a document entitled *Foundations for Learning Provincial Improvement Plan 2009–2011* which articulated the aim of all learners by the end of the 2011 academic year in the foundation and intermediate to increase proficiency in numeracy and mathematics by 30% to meet the National Department of Education's improvement target of 50% by 2011.

"*For increased monitoring of learner achievement, teachers in Grades 1-6 will administer quarterly district-wide math assessment and use the data from the assessments to adjust/revise the action plan"* (GDE, 2009).

The plan included targets, indicators, district strategies, responsibility of schools and teachers, monitoring plans and on-going professional development. In Chapter 6 commentary is provided that indicates the teachers in this study held varying views on the usefulness of the FFL on their design of assessments, while acknowledging that it did provide an impetus for some districts in the province to formulate district common papers especially for under-resourced schools (such as those in the township areas).

The education sector plan

The policy discourse on assessment in South Africa stipulates the roles of teachers and the national ministry in assessment for both formative learning and in the monitoring of learning.

> *"Teachers have the overall responsibility to assess the progress of learners in achieving the expected outcomes and the national and provincial departments of education are accountable for the management of the assessment programmes" (DoE, 2005b, p.7).*

In 2010, the DBE drafted a new broad sector plan. Within this sector plan, regular and standardised testing was regarded as an important intervention to annually measure progress on learner achievement towards a desired target of 60% learner competency by 2014. According to the DBE (2011) the Action Plan specified that the ANA was to be used as part of a testing programme requiring all schools in the country to conduct the same grade-specific Language and Mathematics tests for Grades 1 to 6 and Grade 9. Like the FFL campaign, the standardising of assessment as a tool to 'improve performance' became an explicit goal. In line with similar standardised assessment programmes such as SACMEQ and PISA, the first 3 of the 27 goals included in the education sector plan set to address improvements in Literacy and Numeracy at the key transitional grades, viz. Grades 3, 6 and 9, and progress in this regard was to be measured through the ANA:

> *ANA was put in place by the DBE as a strategy to annually measure progress in learner achievement towards the 2014 target of ensuring that at least 60% of learners achieve acceptable levels in literacy and numeracy. ANA is one of*

> *the initiatives that form the backbone of the DBE's Action Plan to 2014: Towards the Realisation of Schooling 2025 (DBE, 2010).*

The DBE articulated the target of 60% as the percentage of learners they wanted to be achieving the required competencies at the expected grade level by 2014 and noted that the ANA tests were designed to provide information that would help to improve and deliver quality education as part of the State priorities and goals stipulated in Action Plan (DBE, 2010). The State policies suggested that ANA would provide diagnostic information on learner competencies as well while also serving the purpose of monitoring both the academic progress of learners at key stages and schools towards desired targets of the State. Overall in policy imperatives and rhetoric around SNA, there is mention of multiple strands of Morgan's assessment discourses, with the tension that diagnostic information on learner competence is set within the confines of a standardised curriculum, rather than working upwards from learners' existing competences. The support for comparison within this standardisation was also explicit, and viewed as helpful for broad participation of role players.

> *Poor ANA results can also alert districts to the fact that certain things are not working as they should. ANA results will enable districts, parents and schools to have a standard source of information to determine which schools, learners and teachers are most urgently in need of support (DBE, 2011a).*

In Chapter 2, I indicated that to understand the motives of SNA it is necessary to understand its key purposes. As Postlethwaite and Kellaghan (2008) have described the purposes of SNA are multifaceted. In the South African context, the motives of RNCS, the FFL Campaign and the *Action Plan 2014* were structured to enable the state to be the key driver in regulating and standardising curriculum implementation and improving learner performance. Referring back to Morgan's frame, the three policy imperatives described above suggested that multiple discourses were at play with overlaps between characteristics of psychological, curriculum implementation and curriculum standards strands, but with more emphasis placed on monitoring and enforcing the policy goal of curriculum implementation than on establishing the psychological needs of individual learners and setting curriculum standards (and

targets).

5.2.2 The goals of SA-SNA

From the above motives, the State and the DBE identified goals of SNA as critical for monitoring and improving the quality of education in South Africa. The DBE indicated multiple SNA goals in its guideline document on the interpretation and use of ANA results (DBE, 2011). The 2011 guideline document was a preferred source as the goals tied to ANA were explicitly listed, while noting that similar SNA goals had been part of the advocated assessment landscape and media documentation since the launch of the ANA as part of the FFL in 2008 by the Minister of Education. Expanding on these broad motives that had been advocated, the following SNA goals (p.5) aimed at serving multiple communities:

a) Provide the DBE with important information that will help to identify areas where urgent attention is required in order to help improve learning success levels of learners;
b) Assist provincial departments, including district offices, to make informed decisions about which schools require urgent attention in terms of providing necessary resources and support to improve learner performance in these subjects/learning areas;
c) Inform government and the South African public as to how well the schools are serving the country's children where it matters most, namely, in the attainment of functional literacy and numeracy skills that will enable them to study successfully in all subjects and to compete equitably in the labour market;
d) Provide teachers with essential data about the baseline literacy/language and numeracy/mathematics capabilities of learners at the beginning of each grade and thereby help them make informed decisions when planning the year's programme;
e) Inform individual teachers about how close they are to realizing the target goals they seek to attain through their teaching, and inspire them to realign their teaching strategies towards accomplishing such goals.
f) Provide parents with a better picture of the levels of learner performance in the school so that parents are better informed when they become involved in

efforts to improve performance, for instance through decision-making in the school governing body and support to learners in the home.

g) Provide an appropriate benchmark for teachers in the development of assessment tasks that form part of their SBA programme.

h) Assist school management teams to select and implement school-based interventions for improving learner performance in Languages and Mathematics

The above list indicates that more systemic accountability goals than psychological learner goals were stated in the SA-SNA context. These ANA goals can be traced back to the SNA goals (identified in Chapter 2) of Postlethwaite & Kellaghan (2008). For example, the first bullet points towards a systemically diagnostic goal of providing data on learner achievement. In the second bullet, the goal points towards an accountability mechanism to rate schools. Point c) can be linked to the SNA goal of providing data on social and economic objectives, and school-related variables believed to relate to these objectives.

Points d), e), g) and h) above relate to the SNA goal of providing information for policy decisions, particularly at school level. These have implications for teachers and the school management where there is a need to: provide information to teachers through *"credible testing practices"* (DBE, 2011a), be informative about whether targets are being reached with classroom assessment instruments and tasks, appropriately benchmark assessment artifacts utilised by teachers and manage school based interventions. Point e) in particular draws attention to teachers using the information provided by ANA to realigning their teaching strategies towards realising the goals expected of them in delivering the curriculum to learners. Point g) is an explicit comment that SNA ought to guide and influence SBA content selections. This point is central to this study in analysing the SBA programme of teachers to establish the extent of SNA and the ANA being used as the benchmark for the standards of questions to be included in their tasks. Point f) can be linked to the SNA goal of providing data on curriculum reform and giving parents an informed picture of the school's performance in light of the curriculum measures and support to learners being driven by the state. Using Morgan's frame, the SA-SNA goals were analysed and located in specific discourses. This was important to understand the

influences on design rules.

These SA-SNA goals provided the context to explore the operational conditions of assessment activity in terms of design features and advocated rules and the extent to which these advocated rules figured within both the SNA and ANA artefacts and within teachers' assessment artefacts in their SBA.

5.2.3 Advocated design features in SA-SNA

As discussed in Chapter 2, the ANA mathematics tests were based on standardised frameworks similar to that used in international and large scale assessment designs. These formed the "blueprint" for the ANA test developers to design and select suitable questions appropriate for Grade 6 mathematics (DBE, 2011b). The term 'blueprint' usually refers to the specifications of the criteria that final test items must meet, including the proportion of items to address each aspect of a curriculum domain, test length, item format, and any other criteria or constraints regarding test development (Anderson and Morgan, 2008). In Chapter 2, attention was drawn to the make-up of test blueprints by Linn and Miller (2004) and Withers (2005).

In Chapter 2, a specimen framework for systemic evaluations used by the DBE (e.g. the 2004 SE study) was provided with the selected criteria and indicators. Each of the indicators discussed then serves as reference here. Similar fields were listed in the 2010 ANA test framework (DBE, 2011b) for Grade 6 Mathematics. These involved: a) Learning Outcome (LO), b) Assessment standard (AS) focus, c) Skills, d) Number of questions, e) Cognitive level, f) Difficulty level and g) Weighting. **Table 5.3** below shows an extract from the 2010 ANA test framework for LO 2. A key for understanding the abbreviations listed is included below the table.

Table 5.3: Extract of the ANA 2010 Grade 6 Mathematics test framework

Leaning Outcome	Assessment standard focus	Skills/competencies assessed Testing whether the learner is able to…	Number of questions	Cognitive Level	Difficulty level	Weighting (%) for LO
LO 2	Patterns	Investigate and extend geometric patterns looking for a relationship.	1	N	D	15

Patterns, functions and algebra		Investigate and extend numeric patterns not limited to sequences involving constant difference or ratio.	1	A	M	
	Equivalent representations	Write number sentences to describe a problem situation	1	A	M	
	Equations	Solves or completes number sentences (by trail and improvement, inspection, etc.)	1 (with sub-questions)	K	E	

Key: K – knowledge; A – application; N – non-routine problem solving; E – easy; M –moderate; D – difficult; MCQ – multiple choice question; SA – short answer.
Source: 2010 ANA test framework

As with the blueprints suggested by Linn and Miller (2004) and Withers (2005), the format of the ANA framework was a matrix listing for each question (vertical dimension) in the test and item description (horizontal dimension). In the vertical dimension the question number and a brief description of the assessment technique involved was listed. Against this the horizontal dimension provided the question (or item) characteristics which included the intended LO, AS, cognitive level, difficulty level, item type and maximum score. The LO and the AS were drawn directly from the RNCS for Grade 6 Mathematics.

The cognitive and difficulty levels had a 3–way classification. Cognitive levels included: knowledge of basic concepts (K), the application of concepts (A) and non–routine problem solving (N). Difficulty levels included easy (E), moderate (M) and difficult (D). The weighting referred to the expected coverage of the LO as indicated in the RNCS Teacher's Guide for the development of learning programmes (DoE, 2003a). A list of rules is developed according to the purpose of the assessment (Withers, 2005). Additional criteria on the difficulty levels and cognitive demand of items make the rules more explicit. In the ANA reports, the DBE further indicated the intended spread of difficulty and cognitive levels, items had to tap into (DBE, 2011d). This suggested that test developers designed tests according to the percentages indicated in **Tables 5.4** and **5.5**.

Table 5.4: The difficulty levels

Difficulty levels *	**Easy E**	**Moderate M**	**Difficult D**
%	20	60	20

Table 5.5: The cognitive levels

Cognitive levels **	Knowledge of basic concepts (K)	Application of concepts (A)	Non-routine problem solving (N)
%	20	60	20

The above classification tables suggest that the ANA tests were designed to comprise of questions where the majority (60%) are of moderate difficulty and involve the application of concepts. The rationale of the DBE methodology was to "enable a moderately performing learner to succeed without denying outstanding learners the opportunity to demonstrate their performance" (DBE, 2011d, p.12). The design methodology included piloting of draft tests and analysis to check whether assumptions on difficulty and cognitive levels were correct. Thereafter, the final selection of items was decided by subject experts on their appropriateness and alignment to the frameworks. In this study, it was decided that classifications will be analysed more from a test design perspective which indicate the desired expectations and less from the test performance perspective which in the published results reflect that the actual achieved classifications are contrary to above tables.

5.3 EMT analysis on the design rules

The purpose of the EMT analysis is to show up overlaps and tensions that exist between the motives, goals and rules in the SA-SNA activity system. The analysis involved: a) constituting the SNA activity system in terms of the EMT model, b) an analysis of mediated relationships within the activity system and c) the identification of overlaps and tensions.

5.3.1 Constituting the SA–SNA activity system

The SA-SNA activity system is represented in **Figure 5.1**. The focal tools in the SA-SNA activity system for this analysis were the ANA test papers. Within broader policy collectives, motives for standardised assessment activities in Mathematics could be identified. Specific goals on assessment were indicated in the ANA guideline document and these could be linked to SNA goals suggested by Postlethwaite and Kellaghan (2008). The goals of the SNA were identified to be a standardised measure of how well learners are performing in mathematics against learning outcomes and assessment standards outlined in the national curriculum. The

intended subject for design was test developers and policy experts appointed by the DBE. The administrative work in SNA involved a broad community including policy makers, administrators, and teachers at school level but the design work is done mainly by a community of test developers aided and supported by statisticians, psychometricians, curriculum experts, moderators and language editors. Designers needed to ensure that appropriate content was included in the tool with the view, or intended rule, that teachers had to ensure that their learners were exposed to, and assessed on, curriculum concepts included in the ANA (division of labour). Within the SA-SNA system, there was a clear intent on promoting alignment of SNA to the SBA of teachers.

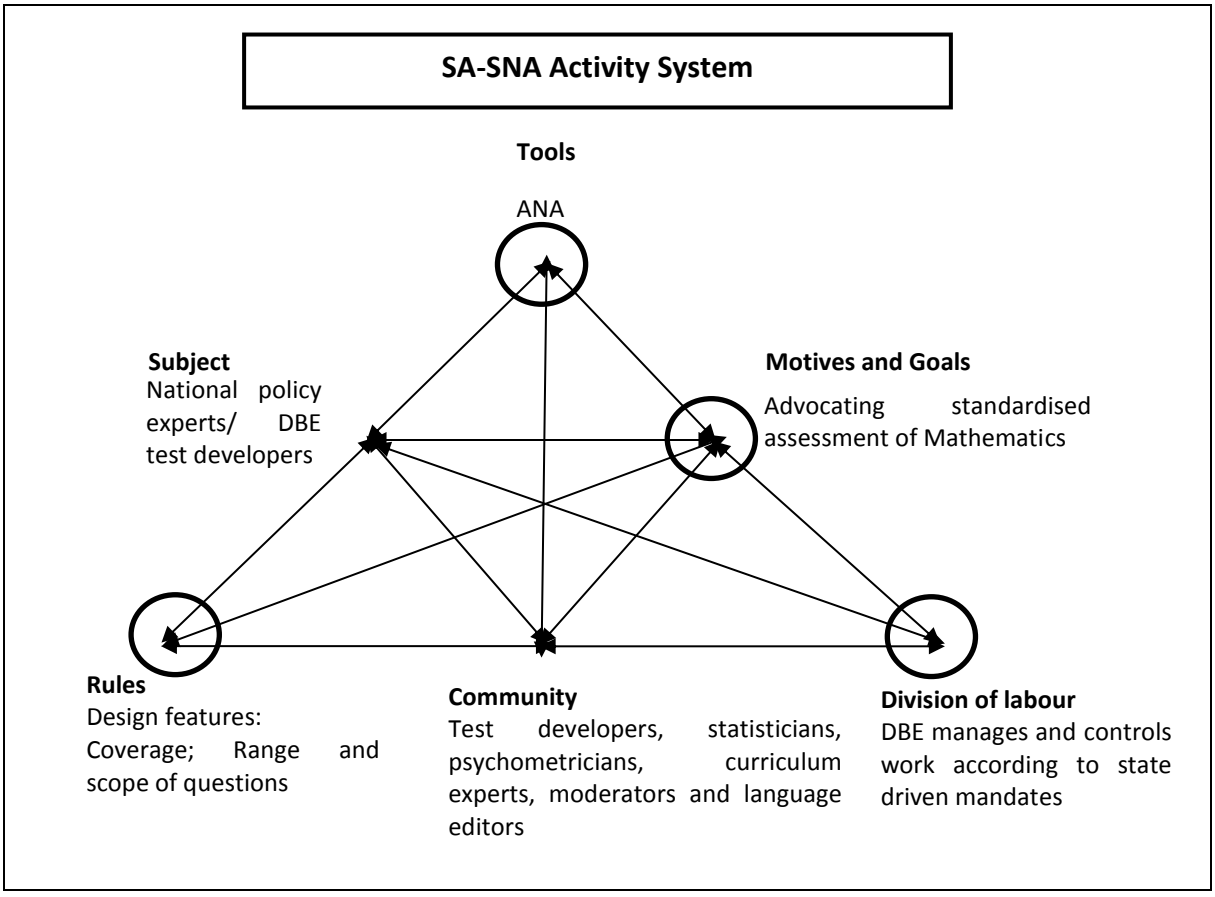


Figure 5.1: EMT model for the SA–SNA Activity System

A critical element in the SA-SNA system in this study was the rules that provided the expected norms for design. National assessments are generally governed by clear specifications which become the rules to generate the assessment tools. In Chapter 2 guidelines by Anderson (2005) for formulating a good blueprint or test specification were listed. In the SA-SNA activity system, the rules considered were the design

features on coverage and the range and scope of questions.

5.3.2 Mediated relationships within the EMT

The next steps were to examine the rules and mediated relationships in the EMT model, and then discuss overlaps and tensions. This involved looking at how design rules are mediated with identified elements within the activity system,

Rules and the mediation of multiple goals

> *There should be regular, external, systemic and national assessment of Mathematics. The analysis of the tests should be used to diagnose areas of focus for interventions and teacher support (DBE, 2009).*

Using the assessment frame supplied by Morgan (2000), discussed in Chapter 2, the goals that the DBE are seeking through SNA with the ANA as the key artefact could be located in the following discourses: psychological, curriculum reform and curriculum standards. In Chapter 2, the point was also made that the RNCS had multiple and contested goals (Chisholm, 2005) due to the input and representation of various stakeholders and within this socially generated product, assessment activities conducted within it would not be immune from such influences and contests among the stakeholder groups. The goals of SA-SNA thus reflected varied social influences drawn from a wide community that included teachers, district officials, policy makers and parents. **Table 5.6** shows how the multiple goals of ANA (DBE, 2011a) are located within different assessment discourses. With each goal, an emphasis on how assessment information will be used is indicated.

Table 5.6: SA-SNA goals located within different assessment discourses

Goal	Psychological	Curriculum Implementation	Curriculum Standards
1. Provide the DBE with important information that will help the Department to identify areas where urgent attention is required in order to help improve learning success levels of learners.	Generate item analysis based on individual learner responses.		
2. Assist provincial departments, including district offices, to make			Produce provincial and

informed decisions about which schools require urgent attention to provide necessary resources and support.			district reports focusing on higher achievement
3. Inform government and the South African public as to how well the schools are serving the country's children in the attainment of functional literacy and numeracy skills that will enable them to study successfully in all subjects and to compete equitably in the labour market.			Promote System-wide outcomes with better individual opportunities as the national economy will improve
4. Provide teachers with essential data about the baseline capabilities of learners at the beginning of each grade and thereby help them make informed decisions when planning the year's programme;	Produce baseline assessment reports so that teaching can be matched to learning needs		
5. Inform individual teachers about how close they are to realizing the target goals and inspire them to realign their teaching strategies towards accomplishing such goals.		Promote teachers to implement changes in curriculum and teaching methods	Promote target setting so that teachers can adopt strategies that will lead to higher outcomes
6. Provide parents with a better picture of the levels of learner performance in the school so that parents are better informed to provide support.	Provide parents with learner information to know and support their children.		
7. Provide an appropriate benchmark for teachers in the development of assessment tasks that form part of their SBA programme.		Promote authentic assessment that matches the values of the desired curriculum	Promote assessments that are normative and challenging.
8. Assist school management teams to select and implement school-based interventions for improving learner performance			Promote interventions that will produce higher achievement

(Source: DBE 2011: A guideline for the interpretation and use of ANA results)

Using the above frame, it is clear that the goals of ANA straddle across the three assessment discourses that Morgan (2000) mentioned. The goals listed as 1, 4 and 6 in **Table 5.6** above have a more psychological (and diagnostic) emphasis on learner abilities. Goals 2, 3, 5, 7 and 8 are curriculum standards based where the emphasis is on promoting system wide targets and strategies that will lead to higher outcomes and better opportunities for learners. Goals 5 and 7 could be located in both the curriculum implementation and curriculum standards discourse where teachers would use the ANA information to realign their teaching strategies towards accomplishing set goals and use the ANA tools as a benchmark for developing SBA tasks. More than half of the goals though, are located in the curriculum standards discourse. de Lange (1999) pointed out that a commonly raised concern with large-scale assessment is whether or not such tests are really about measurement of learner achievement. He has noted that on the policy side, performance measures are not really about measurement, but about political communication. In this study the goals of SA-SNA were seen to fit into a context of state driven policy imperatives which straddled different assessment discourses described in terms of Morgan's framework.

Rules and the development of tools

The documented evidence of the development of the ANA suggests that during the stages of the test design, various role players were appointed by the DBE to manage the required processes. The ANA test specification frameworks were compiled by the test developers who were experienced subject experts appointed by the DBE (DBE, 2012b). The ANA tests were based on the RNCS and developed for Mathematics and Language. In SNA, test frameworks form the "blueprint" for test developers to design and select suitable questions appropriate for a specific grade and then use them to construct a time bound test instrument and memoranda. An extract from the 2010 ANA test framework was shown earlier in this chapter (see **Table 5.3**).

The test blueprint for the ANA describes the content domains and skills levels of the actual construct (Mathematics for the sixth grade) to be assessed. The test development team's remit was to ensure that the test did what it was intended to,

that is, to measure accurately the students' knowledge and skills in the areas identified in the test blueprint. It is unclear from the documented evidence published by the DBE whether the ANA test development team applied any formal review of their final instruments to check whether intended item classification matched the actual learner performance. For example, it is not known whether learners found items classified as easy to be easy. The historical low performance of learners in the ANA suggests that there may be a disjuncture between standards set by policy makers and the standards achieved by the learners. From the literature reviewed in Chapter 2, test developers should design national assessments that contain a representative set of tasks covering the knowledge, skills, and strategies needed for the activity or domain being assessed (Bloom, Hastings & Madaus, 1971).

The judgment of curriculum specialists is important for this purpose. Psychological discourse theorists (Morgan, 2000) suggest that tasks should be aligned to the level of competence of the learners being assessed. Often, the tasks in national assessments in developing countries are critiqued for being based on an "idealised view of achievement and fail to take adequate account of students' current level of achievement or of the conditions in which learning takes place" (Anderson and Morgan, 2008). This point has been echoed by Spaull (2015, p. 135) in the South African ANA context:

> *Parents of primary school children lack reliable information on the performance of their children relative to normal benchmarks (like being able to read by 8 years of age), or relative to socio-economically similar schools in the region.*

Considering the low performance of learners in the ANA tests indicated in the published results (DBE, 2011d) there were also possible disparities between test items used by teachers in their SBA and those selected in the ANA tests. This, in turn, may reflect a division of labour within test construction with limited involvement of practicing teachers from a broad range of SA schooling contexts.

The DBE suggests that their pilot study processes assist test developers to select items for the final tests from those items with desirable levels of difficulty and which perform well as discriminator items (DBE, 2011d), i.e. items that discriminate

performance (using Rasch analysis) between high and low ability learners. The analyses of the pilot studies however were not published in the DBE documentation to establish whether the results of learners matched the desired design and purposes listed above or the extent to which the Rasch analysis was applied in the final selection of test items. Again, the recurring low mathematics results of learners in the ANA between 2008 and 2011 suggest that while the statistical item information is considered important, the final decision may be based primarily on the perceived importance of the item in the curriculum. The ANA results support a view that SNA rules are driven by a curriculum standards/implementation approach, rather than a psychological approach in Morgan's (2000) terms.

5.3.3 Overlaps and tensions

The EMT was examined to identify overlaps and tensions (contradictions) and in the SA-SNA activity system. The analysis pointed out the following primary and secondary contradictions in the SA-SNA context.

Primary contradictions:

A primary contradiction lies in the overlapping goals of the SA-SNA system. The location of the goals of ANA in different discourses leads to questions of whether the design of a single time bound standardised tool can deliver on such a broad variety of goals. From the analysis of the listed goals of SA-SNA, the varying educational goals straddle both psychological and curriculum discourses while mediated relationships within the system show that SNA mandates are more curriculum-oriented than psychological. At certain points of implementation in classroom practice these discourses are likely to be in conflict with each other. A standardised time bound single assessment such as the ANA cannot give a teacher substantive diagnostic information about learner difficulties and serve as a benchmark on the curriculum standard to be attained. Noting the design features of ANA, it only covers a limited range of skills and content (the analysis on curriculum coverage is provided later in this chapter) and DBE reports have utilised the results mainly for system-wide reporting of learning outcomes.

Secondary contradictions

Secondary contradictions were identified between goals and community, goals and tools, and goals, community and rules.

a) Goals and community

The goals of SA-SNA indicate that different communities are targeted for utilisation of information. The goals go beyond just assessing learner achievement on curriculum competencies. There are expectations for parents, school managers and district officials to use the information to structure support and intervention programmes. In this regard the ANA tool is regarded as a systemic planning device which in a short time period must supply information to various stakeholders. National assessment studies such as TIMSS, PIRLS and SACMEQ have almost a 3–4 year cycle to report broader systemic information to a smaller set of stakeholders. This suggests possibilities for time limitations in annual cycle tests to compromise the supply of in-depth and useful information to the community.

b) Goals and tools

The purposes and goals of standardised assessments are usually based on information about student achievement, but can also include information about the factors in the home and in the school that might affect achievement. This has been evident in the international assessments (e.g. TIMSS, PISA and SACMEQ) mentioned in Chapter 2.

c) Goals, community and rules

One among the multiple goals was for the ANA to be an appropriate benchmark for teachers in the development of assessment tasks that form part of their SBA programme. While assessment guidelines have been widely distributed to schools indicating the expected coverage, there has been no distribution of test design features to teachers and limited training on how, at school level, they can model their own assessments on the ANA. For this to happen, explicit test frameworks need to be developed not only for test developers but also for teachers for their SBA programmes. The likelihood here, as reported in the literature on high-stakes assessment (Anderson and Morgan, 2008) is for a “backwash” effect from assessment artefact (e.g. the ANA) from the SNA system into the SBA system. Thus, if there are limitations in the SNA artefact, or slippage in relation to the advocated design features, these are likely to create contradictions with SBA artefact designs. This leads to my analysis, in the next part of this chapter of SA-SNA

artefacts.

5.4 The assessment artefacts

In Chapter 2, design features common to both SNA and SBA artefacts were summarised from literature. These included the coverage of curriculum content, the difficulty level, the cognitive demand of questions, the language demand (amount of text usage) of content and the format of tasks. These assessment features form the basis of the analysis presented in this section.

Drawing on the literature discussed on common design features in SNA and SBA in Chapter 2 and the AT concepts on tools and rules discussed in Chapter 3, the selection of mathematics that was evident in the ANA (based on the 2008/9/10 national papers) is analysed. In this section the analysis is structured on the two themes identified in Chapter 4, namely, the coverage of content and the range and scope of questions.

The key assessment artefacts for SA-SNA were the 2008, 2009 and 2010 ANA tests for Grade 6 Mathematics that was administered nationally in all public schools. For each of these tests, an analysis of mathematical selections was undertaken according to the two themes to gain greater insight into assessment design. This involved identifying for each test item (question and/or sub-question) the item map characteristics: coverage considered in terms of a) the LO, and b) the AS; and range and scope considered in terms of c) the difficulty level, d) the cognitive demand, e) the language demand and f) the item format. An item map was generated for each of the 2008–2010 ANA assessments.

From the item mapping of individual questions, frequency counts were collated for a specific item characteristic (e.g. LO). Each count represented a specific item where learners were required to make a response. For example in the 2008 ANA there were 39 test items (including sub-questions) represented in the test. Each of these test items was linked to a LO according to their representation in the RNCS. Items addressing different LOs were rare and occurred only a few times across the items in all three ANA test papers. Where test items overlapped across different LOs, the LO that comprised the primary skill to be assessed was selected. This count at the item

level broadly reflected the mark allocations as the vast majority of items were awarded one mark across the test papers.

From the frequency counts, I generated summary analysis tables and represented them in percentage form. Summary analysis tables and graphs were then generated within the two themes for each of the item characteristics listed above. Illustrations of these graphs are presented in the discussion that follows.

5.4.1 Mathematical coverage

Mathematical domain and AS coverage were sub-categories chosen to explore the coverage theme in more detail. These sub-categories gave more insight of coverage issues than the designers of the ANA tools may have considered in the selection of mathematical content.

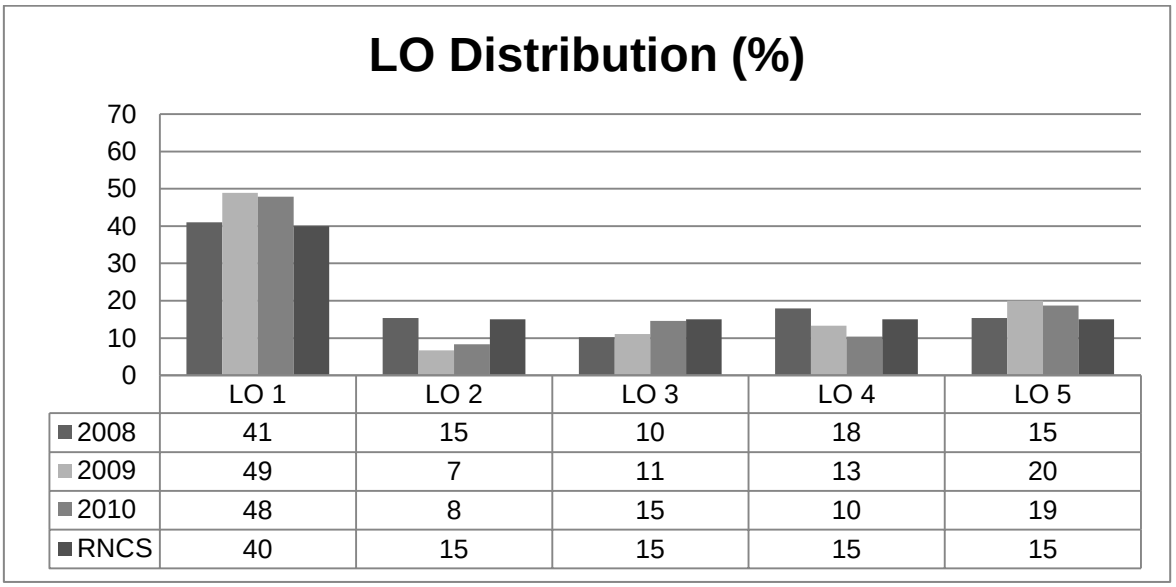
The RNCS clearly stipulates the mathematical domains that learners should be exposed to in assessment tasks (DoE, 2002b). The five mathematical domains were listed in Chapter 2 and have reference here. The expected LO weightings of the RNCS on mathematics content coverage for Intermediate Phase were 40% for LO 1 and 15% for each of LO2, LO3, LO4 and LO5 respectively (DoE, 2003a). These weightings were carried through into the FFL milestones coverage for each term. The LO distribution across the three years 2008–2010 ANA papers could thus be compared against these expected weightings. Beginning with aggregated counts of LOs across all items in the papers produced the summary shown in **Table 5.7**.

Table 5.7: Aggregated counts of LOs

ANA	2008	2009	2010
LO 1: Numbers, Operations and Relationships	16	22	23
LO 2: Patterns, Functions and Algebra	6	3	4
LO 3: Space and Shape (Geometry)	4	5	7
LO 4: Measurement	7	6	5
LO 5: Data Handling	6	9	9
Total count of test items	39	45	48

In **Table 5.7**, the LO distribution counts for each of the three ANAs are represented. For example in the 2008 ANA, of the 39 test items, there were 16 counts of LO 1, 6 counts of LO 2, 4 counts of LO3, 7 counts of LO 4, and 6 counts of LO 5 represented

in the test. Hence, the LO 1 distribution percentage was calculated as 41% (i.e. 16 out of 39). **Figure 5.2** illustrates the LO distribution (%) for the three ANAs and the expected LO weightings of the RNCS.



	LO 1	LO 2	LO 3	LO 4	LO 5
2008	41	15	10	18	15
2009	49	7	11	13	20
2010	48	8	15	10	19
RNCS	40	15	15	15	15

Figure 5.2: LO % distribution 2008–2010

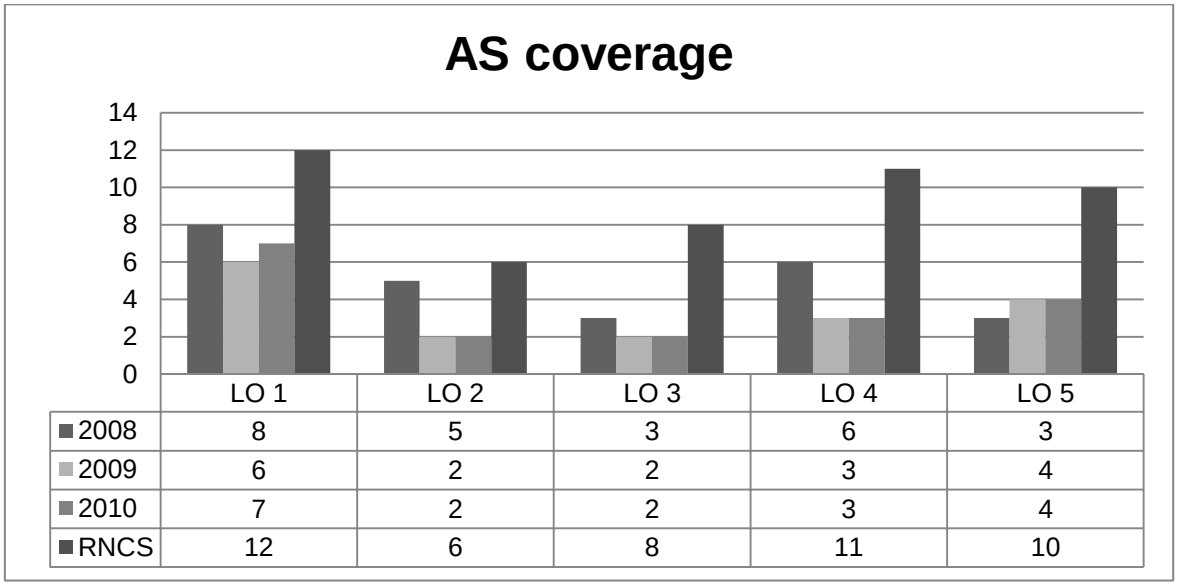
In **Figure 5.2**, LO1 had the highest weighting across all three years, with notably higher than the expected weighting of 40% in 2009 and 2010. Of the 45 items included in the 2009 test, approximately 49% of the items tapped into LO1 skills. The remainder of the items spanned across the four other LOs (LO2 = 7%, LO3 = 11% LO4 = 13% and LO5 = 20%). Across the three years, the distribution reflected coverage of all five LOs but the closest match to the expected weightings occurred in 2008.

The 2009 and 2010 tests, when compared to the expected weightings of the RNCS, allocated over-emphases to LO1 and LO5. For example in the 2009 test there was an over emphasis in LO1 by 9% and in LO 5 by 5% and under-estimates in LO 2 by 8% and LO3 by 4%. The emphasis on LO 5 exceeded the recommended weighting which could have been reduced to assess more skills in LO 2 and LO 3. The lowered weighting of LO2 is particularly problematic given that at secondary level and in Grade 12, LO2 has a high weighting overall.

Coverage of the ANA tests also had to be looked at in terms of the ASs and the nature of the mathematical skills assessed within each LO across the three years. In

Chapter 2, I noted that ASs describe the level at which learners should demonstrate their achievement of LOs and the ways (breadth and depth) of demonstrating this achievement. In this study, I looked at breadth across AS in terms of how many different ASs were covered in the tests. Each different AS was considered as a single count. These single counts could then be compared to the number of ASs per LO indicated in the RNCS. The depth of an AS referred to the assessed skill and concept that was covered, and was considered in terms of DL, CD and LF.

Figure 5.3 summarises the coverage of ASs in the 2008, 2009 and 2010 ANA tests. Only unique, non-repetitive counts of ASs were considered. For example, the 2008 ANA had test items that covered 8 unique ASs from LO 1 and 6 individual ASs from LO 4. In **Figure 5.3**, this unique spread of ASs per LO for each of the three ANA tests is compared with the total number of ASs indicated for each LO in the RNCS. The total number of ASs for the RNCS Grade 6 Mathematics was listed in Chapter 2, and have reference here.



	LO 1	LO 2	LO 3	LO 4	LO 5
2008	8	5	3	6	3
2009	6	2	2	3	4
2010	7	2	2	3	4
RNCS	12	6	8	11	10

Figure 5.3: Coverage of assessment standards in the ANA tests

In **Figure 5.3**, there are 47 assessment standards spread across the five LOs in the RNCS. While it is generally expected that time limited systemic assessment will not cover all the assessment standards in the curriculum, my interest in the above figures relates to variations in the breadth of coverage. The 2008 ANA had the widest range of ASs assessed (25) compared to 2009 (17) and 2010 (18). Looking across all five LOs, in the 2009 and the 2010 ANA tests, the shortfall was most

striking in LO 2, LO 3, LO 4 and LO 5, where there was less than 50% coverage of the total number of ASs. The implication of this is that several ASs would not have been assessed in the ANA tests. The reduced item coverage in certain ASs (e.g. interpreting information from graphs) results in limited diagnostic and formative information that can be extracted from the ANA tests, while also being potentially limiting in terms of the "backwash" effect into SBA. This is compounded by DBE reporting formats that have not made explicit claims about how learners have performed in relation to specific LOs and ASs.

Overall, there is a degree of overlap across the three years in terms of the coverage of different ASs within the five LOs with coverage of ASs in 2009 and 2010 being notably similar. In all three years there was coverage of ASs across all five content domains (LOs). Although the 2010 test had the highest number of test items (48), the 2008 ANA had the widest range of ASs assessed (25). In the next section a closer look at the range and scope of these questions is discussed.

5.4.2 Range and scope of questions

With reference to the literature reviewed in Chapter 2, the range and scope of test items included in the ANA were considered in terms of: difficulty level (DL), cognitive demand (CD), language demand (LD) and the item format (IF).

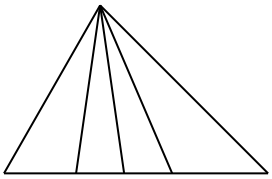
Difficulty level

The descriptors of difficulty included in the ANA frameworks had a 3-way classification, namely, easy, moderate and difficult. The DBE documentation did not provide detailed descriptors for these classifications with inferences to the items, being subject to a pilot study and checked for their appropriateness and alignment to the desirable levels of difficulty listed in the frameworks. In this study, ways of explaining the 3–way classification of easy, moderate and difficult are drawn from Leong's (2006) frame of linking difficulty level to content knowledge dimensions based on curriculum grade specifications (explained in Chapter 2). Leong's argument that multiple knowledge dimensions coming together make an item harder for a learner to respond to, are used in the SA-SNA context against the grade specific requirements of the RNCS. Using Leong's (2006) content difficulty frame

based and Anderson's (2005) explanation of knowledge dimension elements in his LTA taxonomy allowed for a composite theoretical description of easy, moderate and difficult items in the ANA tests. **Table 5.8** provides the difficulty level (DL) descriptors used in the analysis.

Table 5.8: Difficulty level descriptors

DL descriptor	Description	Test item exemplar
Easy:	Items are structured on basic and factual knowledge elements that learners should have encountered at lower grade levels according to the curriculum. They are viewed as familiar to learners on the basis that they should have had the opportunity to learn them in previous as well as current grades. The number range used in the item is below the expected grade level of the learner as listed in the RNCS.	**Item 1** 4. For each number write the value of the underlined digit. 4.1 3 50**3** _____ 4.2 **3** 503 _____ *The above test item was extracted from the 2009 ANA but the same item also featured in the 2008 and 2010 ANA.*
Moderate:	Test items are based on knowledge elements at the appropriate grade level of the curriculum and are those that are central to the core curriculum These will include items that require conceptual knowledge (knowledge of classifications and categories, principles and generalisations, and theories, models and structures and/or procedural knowledge (knowledge of methods, techniques, algorithms and skills) elements pegged at the current grade level.	**Item 2** 10. Calculate and write your answer as a mixed number. 10.1 $8\frac{3}{10} - 4\frac{1}{5}$ 10.2 $5\frac{1}{2} + 3\frac{1}{8} + \frac{3}{4}$ (4) *Test item extracted from the 2010 ANA test.*
Difficult:	Items are based on advanced knowledge elements and are usually those that should be covered more adequately at more advanced grades and hence are peripheral to the core curriculum. Learners may not have had sufficient opportunities to learn these elements in their current grade. These	**Item 3** 11. The figure below is made up of triangles of different sizes:

	knowledge elements are likely to be difficult for most of the learners. Test items that assess learners on two or more knowledge elements or combines knowledge elements that are seldom combined (Leong, 2006).	 How many triangles are there in this figure? _____ (2) *Test item extracted from the 2010 ANA.*

Examples of items for the three DL levels are shown in the above table. An easy item was one that a learner in the grade assessed should find easy to do on the basis that it is pitched at a lower grade level. Item 1 was classified as easy in the above table since it was pitched at the Grade 4 level. In the RNCS, grade competencies go hand-in-hand with the extension of number range. For example in this particular AS of the curriculum, the learner must show the following grade competencies (DoE, 2002b):

a) Grade 4: *Recognises the place value of digits in whole numbers to at least 4-digit numbers*.

b) Grade 5: *Recognises the place value of digits in whole numbers to at least 6-digit numbers*.

c) Grade 6: *Recognises the place value of digits in whole numbers to at least 9-digit numbers*.

An item of moderate difficulty was pitched at the appropriate level of the assessed grade with the understanding that a learner in Grade 6 should not find it difficult to do or answer. Again the moderate DL descriptor could be matched against the following grade competencies. The AS for this item was as follows (DoE, 2002b):

a) Grade 4: *Addition of common fractions in context.*

b) Grade 5: *Addition and subtraction of common fractions with the same denominator*.

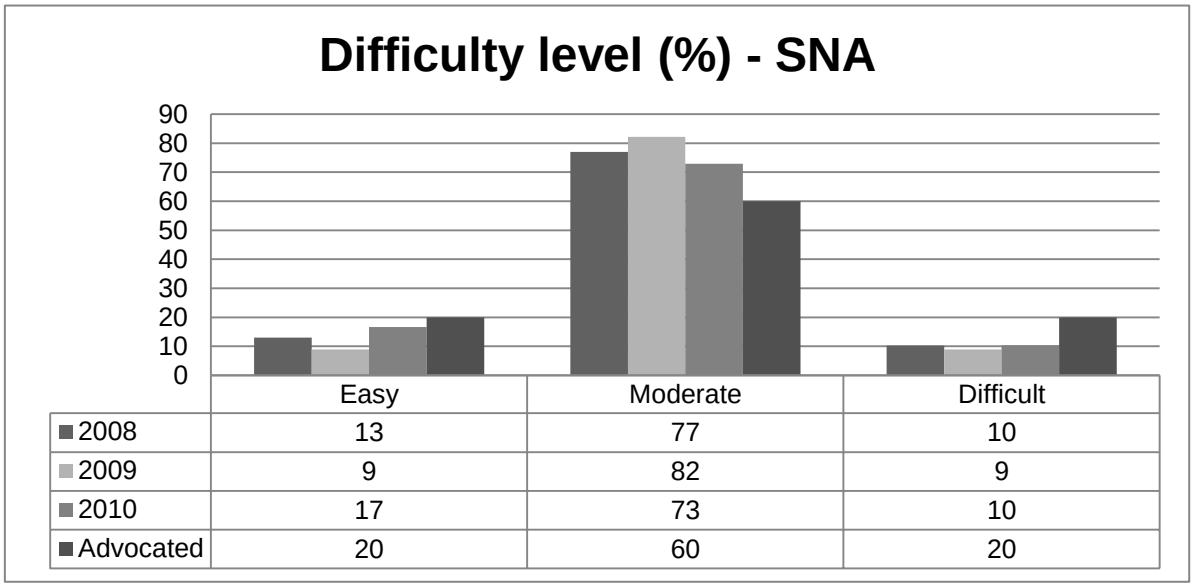
c) Grade 6: *Addition and subtraction of common fractions with denominators which are multiples of each other*.

Item 2 in the above table could therefore be classified as having a moderate DL.

A difficult item was considered to be pitched at a higher level than the assessed grade. Item 3 in the above table was considered difficult due to being ‘unusual’, in

that “composite” triangles have to be counted. Learners had to use grouping techniques, identifying shapes within larger shapes and their knowledge of triangles to find the required solution.

In terms of the DL spread advocated, ANA test designers were directed towards locating DL proportions of test items according to 20% of items being easy, 60% at moderate difficulty and 20% of items should be difficult. The rationale behind these proportions was that the spread provided opportunities for a wide range of learners to demonstrate some competence in the tested skills and knowledge (DBE, 2011d).
In applying the DL frame in the overall item map for the ANA tests of 2008 to 2010, the overall spread of difficulty could be established. For example, in the 2008 ANA test, there were a total of 38 test items, out of which 5 were easy items, 29 items were of moderate difficulty and 4 were difficult items. This translated into DL percentage breakdown of 13% easy, 77% moderate and 10% difficult. **Figure 5.4** shows the percentage spread of difficulty in the ANA tests of 2008–2010 generated from the DL frame.



	Easy	Moderate	Difficult
2008	13	77	10
2009	9	82	9
2010	17	73	10
Advocated	20	60	20

Figure 5.4: Spread of DL % in the ANA tests from 2008–2010

Across the three years the overall spread of difficulty showed similar trends. The number of moderately difficult items was the highest across all three tests and higher than the advocated weighting. The highest percentage of easy items was found in the 2010 ANA (17) and the lowest percentage was in the 2009 test. Across the three tests, the number of difficult items was almost the same. In the 2009 ANA, there

were 4 easy items (9%), 37 moderate (82%) and 4 were difficult (9%). In the 2010 ANA, there were 8 easy items (17%), 35 moderate (73%) and 5 difficult (10%). Across the three years the percentage of easy items was different to the advocated percentage with only the 2010 ANA % close to the expected 20%. The percentage of difficult questions was similar across the three ANAs but significantly (more than 10 percentage points) different to the advocated percentage with the 2009 ANA 11 percentage points lower than the expected 20%.

Cognitive demand

In Chapter 2, cognitive demand (CD) was referred to as the cognitive characteristics of assessment items. In other words the cognitive processing capacity required of a learner in order for him or her to successfully answer the assessment item. The ANA framework represented CD in terms of a spread of: Knowledge of basic concepts (K), Application of routine concepts (A) and non-routine problem solving (N). As a guide, the 2010 ANA mathematics tests cognitive demand spread was expected to have 20% of K-items, 60% A-items and 20% N-items. I noted earlier that DL and CD had a similar spread of proportions for test specifications in SNA documentation.

As with DL, the documentation on the ANA did not provide a description for each of the cognitive levels listed in the framework. However, it was mentioned in Chapter 2 that these classifications do have research precedence in the literature on national assessments and various cognitive classifications from different researchers and institutions of assessment were unpacked. Based on ideas drawn from Anderson's (2005) LTA construct, Webb's (2002) DOK classification system, de Lange's (1999) assessment pyramid, Stein et al (2000) 4-level classification, the TIMSS mathematics blueprint and a blueprint for a middle primary mathematics test in the United Kingdom (Anderson and Morgan, 2008), the following CD frame (see **Table 5.9**) was formulated as an organising tool to locate cognitive characteristics of the ANA test items.

The essential difference between the CD frame indicated below and the ANA test framework was the preference to work with 4-levels rather than the 3 stipulated in the ANA documentation. This allowed for application type problems to be classified

as either routine or complex procedures, a categorization closer to the model used in TIMSS studies. By having separate and differentiated descriptors in the CD frame, there was reduced ambiguity in classifying items. As suggested in Chapter 2, the common view among researchers is that the cognitive demand of assessment tasks at school level should be structured according to hierarchical levels or learning hierarchies. This view is represented in the CD descriptors listed in **Table 5.9**.

Table 5.9: CD descriptors

CD descriptor	Description	Test item exemplar
Knowing basic facts (K)	Items are based on knowledge that require recalling terms, properties and procedures (Anderson, 2005), information, concept or procedure (Webb, 2002). Test items deal with knowledge of facts, representing, recognizing equivalents, recalling mathematical objects and properties (de Lange, 1999). Often these test items involve memorisation (Stein et al, 2000) and relevant facts, properties, procedures would normally be located below the focal grade level.	**Item 1** 21. Answer the following questions. What name can you give to a shape with 21.1 Two pairs of opposite sides' equal and one angle equal to 90°? _____ 21.2 Two pairs of opposite sides' parallel and two pairs of opposite angles equal to 180°? _____ *Test item extracted from the 2009 ANA. Item also featured in the 2008 ANA.*
Applying routine procedures (R)	Items require learners to carry out or select routine operations or procedures expected for that grade. It involves a basic application of a skill or concept (Webb, 2002) It includes understanding of mathematics concepts, principles and structure. Ability to translate elements from one form to another. Ability to read and interpret graphs and diagrams (Anderson, 2005). Ability to solve routine problems. Test items require learners to perform basic procedures, apply standard	**Item 2** 24. The following are the shoe sizes of some Grade 6 learners. 7 8 7 6 5 5 7 4 6 7 8 4 7 5 6 24.1 What is the *mode* of the shoe sizes? _____ 24.2 What is the *median* of the shoe sizes?

	algorithms, and develop technical skills (de Lange, 1999). The procedures used by learners are applied without any complex connections (Stein et al, 2000)	_____ *Test item extracted from the 2010 ANA. The item also featured in the 2009 ANA.*
Using complex procedures (C)	Items require learners to analyze information and make comparisons. It includes strategic thinking using a combination of methods, techniques, algorithms and skills to solve complex problems (Anderson, 2005). Solving these test items requires reasoning, developing a plan or a sequence of steps to approach a problem that may be abstract and complex (Webb, 2002). Although the problems may also be non-routine, they require relatively minor mathematical workings. From the point of view of mathematical language, another aspect at this level is decoding and interpreting symbolic and formal language and understanding its relations to natural language. Items at this level are often placed within a context and engage students in mathematical decision making.	**Item 3** 12. Dr Mololo travels 90 *km* to the hospital. For every 10 *km* that she travels, her car uses 2 litres of petrol. How many litres of petrol does the car use to drive to the hospital? _____ _____ _____ _____ *Test item extracted from the 2010 ANA.*
Solving non-routine problems (N)	At this level, learners are asked to extend their thinking and mathematise situations (recognize and extract the mathematics embedded in the situation and use mathematics to solve the problem). Learners must analyze, interpret, develop their own models and strategies, and make mathematical arguments including proofs and generalizations (de Lange, 1999). The investigation or application to the real world generally requires time to think and process multiple conditions of the problem or task; involving non-routine manipulations (Webb, 2002).	**Item 4** 16. To earn some extra money, Anne makes necklaces to sell to her friends. The one she is making is made up of 3 sets of red beads (✳) and 3 sets of white beads (⬣) and looks like this: ✳⬣⬣⬣✳✳⬣⬣⬣⬣⬣✳✳✳✳⬣⬣⬣⬣⬣⬣⬣ If she continues with this pattern, how many beads of each colour will be in the next set? _____ *Test item extracted from the 2009 ANA.*

Items 1–4 in the above table were selected as examples to illustrate how individual

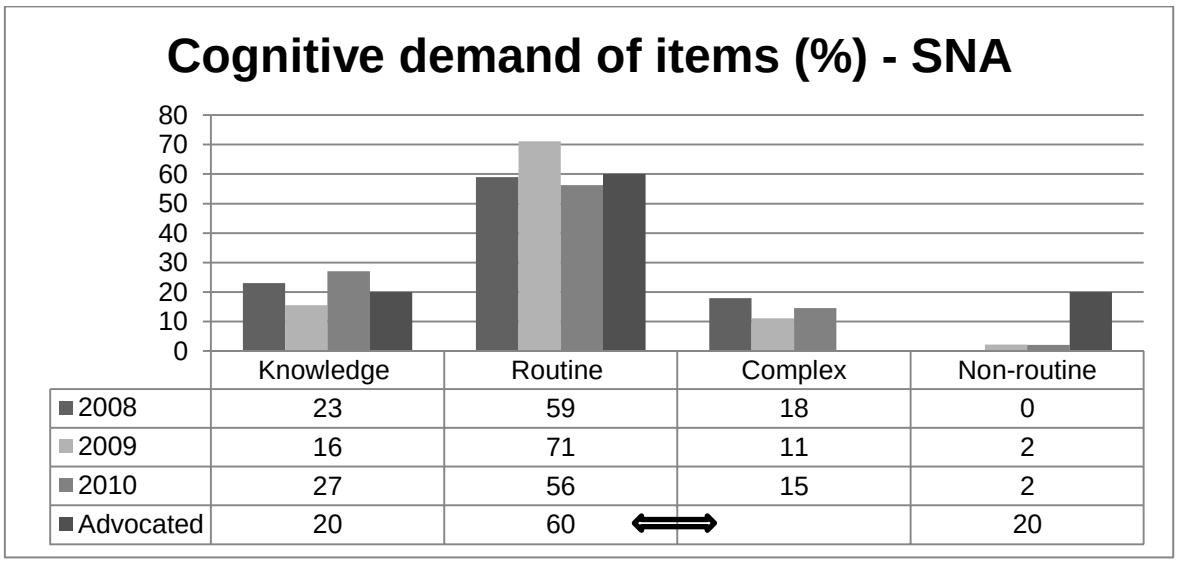
test items were mapped to the 4 CD levels. Item 1 was a knowledge (K) question where learners were expected to know the AS: *basic similarities and differences between rectangles and parallelograms*. It was a knowledge requirement pegged at a Grade 6 level. Item 2 was mapped as a routine (R) application directly linked to the Grade 6 AS where learners were expected to *examine ungrouped numerical data to determine the most frequently occurring score (mode) and the midpoint of the data set in order to describe central tendencies*.

Item 3 was selected to illustrate an item mapped as having a complex (C) cognitive demand. The item was based on the Grade 6 AS where learners had to *solve problems that involved comparing two quantities of different kind* (distance travelled and usage of petrol) The item was regarded as complex as learners were expected to understand the context, analyse the given information on distance travelled and usage of petrol, construct an algorithm to work out the rate of petrol used per kilometer and then work out the solution to the problem posed. The combined use of mathematical reasoning and mechanical skills contributed to regarding this item as complex according to the CD frame.

Item 4 was mapped as non-routine problem solving (N). This item was based on the Grade 6 AS where a learner is expected to *investigate and extend numeric and geometric patterns looking for relationship and rules including patterns represented in physical or diagrammatic form.* The item was considered non-routine because learners were expected to mathematize (de Lange, 1999) the given problem by extracting the relevant mathematical information, investigate the given patterns of red and white beads, analyze the numeric sequences involved for both sets of beads, and then interpret how the patterns can be extended. This item was regarded as going beyond complex problem solving techniques where routine algorithms can be combined as a solution. The learner has to work with two sequences. In the sequence of the red beads, the sequence has elements that can be seen as being doubled as the pattern is extended. There is a constant ratio of 2. In the sequence of white beads, the elements form an increasing odd number pattern. There is a constant difference of 2. Learners would have to work with each sequence independently to solve the problem and the skills involved go beyond the Grade 6 curriculum.

It should be noted that describing CD levels has been acknowledged to be a "somewhat arbitrary activity": there is no clear distinction between different levels, and both higher- and lower-level skills and competencies often play out at different levels (de Lange, 1999). Realistically, in time restricted national assessments such as the ANA, not all cognitive levels of questions can feature in a single assessment tool. Items in the fourth level (non-routine) are difficult to assess in time bound tests (de Lange, 1999) and therefore would be few in number in a standardised assessment such as the ANA. In all three assessments the majority of items were directed towards learners solving questions that were either knowledge or routine. In the 2008 ANA, there were 9 K items, 23 R items, 7 C items and 0 N items. In 2009, the CD spread of items was: 7 K, 32 R, 5 C and 1 N. In 2010 the CD spread was: 13 K, 27 R, 7C and 1 N.

The advocated spread (DBE, 2011d) was knowledge of basic mathematics concepts (20%), application of concepts (60%) and non-routine problem solving (20%). According to the CD frame used in this analysis, application of concepts was split into routine and complex procedures. Therefore the 60% in the advocated row should be seen as split across routine and complex procedures. **Figure 5.5** indicates the CD spread as percentages across the three assessment years.



	Knowledge	Routine	Complex	Non-routine
2008	23	59	18	0
2009	16	71	11	2
2010	27	56	15	2
Advocated	20	60	⟺	20

Figure 5.5: CD spread for ANA 2008–2010

In all three assessments, the CD level with the most number of items was routine

applications. The percentage of routine items in the 2009 ANA (71%) was significantly higher than the other two assessments but had the lowest percentage of complex items. The combined percentages for routine and complex procedures were significantly higher than the advocated 60% and in 2009 the combined percentage of routine and complex problem solving was as high as 82%. The 2008 ANA had the highest percentage of complex questions. The highest percentage of knowledge items featured in the 2010 ANA but the 2008 ANA was closest to the expected 20%. There were also significant differences between the percentage of non-routine questions in the ANA tests and the expected 20%. Across all three years, the percentage of non-routine questions was low and significantly lower than the advocated percentage. In the 2008 ANA, there were no items that could be considered non-routine in terms of the CD frame used. This suggested that while the form of the ANA assessment was driven by a curriculum standards orientation, the selection of content for the ANA appears to diverge from this framework within the SA-SNA context. Put simply, there is 'slippage' from the advocated breakdown of CD as a design feature in the SNA assessment system itself, prior to even considering gaps between what is advocated in SNA and enacted in the SBA activity system.

Language demand

Before an appropriate response can be made learners have to read through and interpret information/instructions given in the stimulus material of the item. Instructions refer to the actions that the designers require the learners to undertake. For example in a multiple choice question the instruction could be: 'Circle the letter of the correct answer.' According to Withers (2005) the stimulus material could be informative (e.g. a passage or diagram in a mathematics test which forms the basis of the solution) or it could be directive which could be a statement of a problem to be solved (e.g. what is the Area of the shaded part of the figure?). In this study, the language demand (LD) of an item referred to the amount and nature of text in the instructions and stimulus material. Research by Leong (2006), Howie (2002), de Lange (1999) and international assessment studies (TIMSS and PIRLS) provided a historical base of criteria to be considered in analysing the role of LD in influencing the construction of items in different assessment tasks. To further unpack the LD used in the ANA tests the following frame (**Table 5.10**) was used.

Table 5.10: LD descriptors

LD descriptor	Description	Test item exemplar
No instructional text (NT) (Computational)	Items that contain no instructional text or no words at all. Words and phrases when used require only simple and straightforward mathematical working (Leong, 2006). There is no instructional verb (e.g. Calculate) used in the item.	**Item 1** 7. 10% of 180 is equal to _____ *Test item extracted from the 2009 ANA*
Low text (LT) (instruction/s is/are short mathematical verb/s)	Test items have instructions that are directed by short mathematical verbs (e.g. Add or Calculate). The Items have language instructions that follow central issues in the stimulus not peripheral or trivial details. The instructional verb directs learners on what they are required to do, it stands alone and does not depend on understanding from a previous item (Howie, 2002). The instruction can include just the verb or a short sentence containing the verb.	**Item 2** 13. Calculate: 18 + (3 x 8) ÷ 6 _____ _____ _____ *Test item extracted from the 2009 ANA*
High text (HT) (word problems)	Items that have more than one sentence in the stem and are often referred to by South African teachers as word problems. The test Items contain information from life situations that are either "real" or "imagined" (de Lange, 1999) The context may be used to build awareness of other subjects (e.g. science) as well as human rights, social, economic and environmental issues (DoE, 2002b).	**Item 3** 15. Last holidays we travelled by car to visit my grandmother who lives 1 248 km away. We stopped to fill up petrol after travelling 569 km. 15.1 How much further from the filling station did we have to travel to reach grandmother's house? *Test item extracted from the 2009 ANA*

The items shown in the above table are examples of the LD levels used in the analysis. Item 1 was mapped as a computational problem having a NT language demand. In this item, there was no instructional verb contained within the stem of the problem. Learners are required to focus on and show an understanding of the

mathematical or symbolic form to generate an appropriate response. Item 2 was mapped as an item with LT language demand containing an instructional mathematical verb. The verb ‘calculate’ in the stem is generally used to direct the learner to the required procedure or steps to take or display in the solution. Howie (2002) suggested that in such items, there is no need for peripheral or trivial information to be included. Other similar examples in ANA tests included verbs such as “Add”, “Round Off”, “Draw”, etc.).

Item 3 was mapped as an example of a word problem considered to have a HT language demand. The language demand required learners to read through text in the initial stem to understand the context and then read through text in the sub-stem to provide an appropriate response. In the South African context, the scenario provided can be reflective of a real world or imagined situation that Grade 6 learners could identify with. The NCS emphasises the application of real world contexts that integrates information and knowledge from other learning areas and social contexts, but an analysis in those terms is not the focus of this study. Instead, given that qualitative diagnostic analysis reports on ANA test items (DBE, 2013) point toward learners not performing well on mathematics items with HT language demand, backed up by findings from the 2011 PIRLS study, my focus on LD was on the potential for language to be used in ways that made items harder to access.

Using the above frame an analysis could be drawn on the LD across the 3 years of ANA assessment papers. **Figure 5.6** shows the LD spread across the ANA tests from 2008–2010.

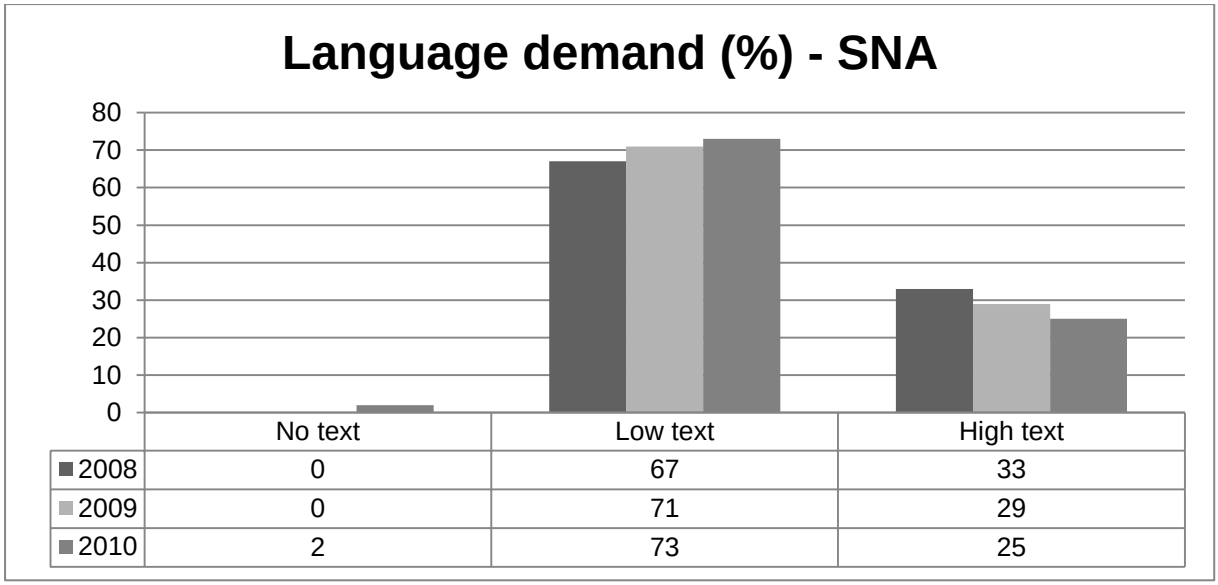


Figure 5.6: LD % spread across for the ANA tests: 2008–2010

Across the 3 years, the LD spread was relatively similar, suggesting directives given to designers during the test development process. Almost no NT computational problems in the strictest sense of the criteria used in the LD frame could be found in the ANA tests of 2008 and 2009. The LD was generally split between high text word problems and those that had low text with an instructional verb telling learners what to do. The percentage spread showed just over two thirds of test items across the three years were located in the LT category with almost 73% of items falling into this category in the 2010 ANA. This implies that across most items the ANA tests provided very specific instructions to learners in short sentences with explicit mathematical verbs. The majority of items on numbers, operations and relationships (LO 1) across the three years had a LT language demand. HT language demand questions featured more in the assessment of items on measurement (e.g. capacity and time) and data handling (e.g. examining ungrouped data, probability and graphs). It was also noted that certain low text questions had additional instructions which could have been left out. For example in the following item extracted from the 2010 ANA, the additional text in the stem includes the words ‘in the following list” to the stem of the question

1. Circle the correct answer

1.1 The multiple of 10 in the following list is.....

102 112 120 153

1.2 The prime number in the following list is.....

19 21 33 39

The above LT item with less text could have been phrased as follows.

1. Circle the correct answer 1.1 The multiple of 10 102 112 120 153 1.2 The prime number 19 21 33 39

The 2011 PIRLS study results showed that the reading levels of primary school learners in South Africa are still below an acceptable level with only 71% of Grade 4 learners able to reach a 'rudimentary' level of reading and 43% of Grade 5 learners cannot read at an appropriate level (Howie et al, 2012). A similar finding could be drawn about the Grade 6 reading levels of learners from the 2006 SACMEQ study. Qualitative diagnostic analysis reports on ANA test items (DBE, 2013) point toward learners not performing well on mathematics items with HT language demand and since almost a third of the questions fell into the HT demand category, it could be a possible explanation for the low overall performance of learners in the ANA mathematics tests. The pertinent point here is that there is a spread between LT and HT on the papers with almost no NT. Prior SA evidence suggests low language and literacy levels in the Intermediate Phase mean that this balance, and particularly the fact that roughly a third of questions are HT is likely to make it harder for children to access the problems. At Grade 6 level, if the instructions are not stipulated in straightforward, unambiguous language the learner may get confused and will not know what do.

Item format

The final assessment feature that was considered within the range and scope of questions was the item format (IF). In paper-and-pen assessments such as the ANA, students respond to a series of questions or prompts. Their written or drawn responses are used as evidence of their level of knowledge, competence, or understanding. In national assessments, there are four basic item formats that are used for learners to show their responses: a) multiple-choice response b) closed constructed response, c) open ended short response and d) essay or extended response (Anderson and Morgan, 2008). In Chapter 2, examples of each type were presented.

Although the ANA framework listed earlier did not include specifications on item type, supporting documentation on the ANA suggested that the designers made use of two specific types of format to elicit learner responses: multiple choice questions (MCQ) and closed short answer questions (CSA). Generally, open ended and extended response type responses did not feature in any of the ANA mathematics tests, and therefore, they were not considered in the analysis. Instead, the analysis was based on the inclusion and nature of MCQ and CSA items.

With the CSA response type items, there is one correct answer that the learner generates. Good short-response items are clear and precise and are developed with scoring guides (Anderson and Morgan, 2008). The scoring guide may reflect partial credit for a CSA item and may differentiate between more comprehensive, precise, or sophisticated responses and incomplete or partially correct responses with better answers awarded a higher score. Descriptions of MCQ and CSA formats and examples from the ANA tests on each format are listed in **Table 5.11**. Using the criteria listed for MCQ and CSA, an IF analysis was carried out across the 3 assessment years.

Item 1 in **Table 5.11** is an example of the typical MCQ item format used in the 2009 ANA. Similar formatting of the MCQ questions featured in 2008 and 2010 ANAs. The number of options across the three assessment years was standardised to 4 choices, in keeping with the international norm. Withers (2005) argues that 3 or 6 option items are avoided as 3 options increases the chances of guessing and 6 options are too complex to handle in the circumstances of a test. In the 2008 ANA, the MCQ questions had letters (a to d) next to the plausible options but in 2009 and 2010 (see item 1), no letters were included alongside the options. In the 2008 and 2009 tests, where an emphasis was required (e.g. prime number), words were underlined.

Table 5.11: IF descriptors

IF descriptor	Description	Test item exemplar
Multiple choice questions (MCQ)	MCQs have one unequivocally "correct" option and several plausible but incorrect options. Multiple-choice items require students to select one of several (usually four) options. Options may be written out or shown as labeled pictures. They may be listed one under the other, shown as a horizontal row, or given in two columns. Students indicate their response by shading a bubble, drawing a ring around an alphabet letter or number, or ticking a box to select a piece of text or a diagram (Anderson and Morgan, 2008).	**Item 1** 6. Circle the number that has the same value as $\frac{3}{10}$. 3%; 30%; 0,03; $\frac{6}{100}$ *Test item extracted from the 2009 ANA*
Closed short answer response questions (CSA)	CSA items have one correct answer that the student generates. In mathematics tests, the learners are often required to answer by a number or value, a mathematical definition and/or property, complete a number sentence, pattern or table, draw and/or interpret a simple graph, make predictions and use formulae accurately. Minor variations in the way the answer is shown are usually acceptable (Anderson and Morgan, 2008).	**Item 2** 9. Solve the following: 9.1 $\frac{5}{6} - \frac{1}{3}$ 9.2 $7\frac{1}{2} + 10\frac{1}{2} + \frac{3}{4}$ *Test item extracted from the 2009 ANA*

Across all 3 years, the majority of test items were of a CSA format with a unique correct answer listed in the scoring (marking) guide. Item 2 is an example from the 2009 ANA test. Here, learners wrote down a response in each of the columns that were numbered based on the figure indicated in the first column. Generally in these types of questions alternative ways of working was accepted but the answer had to be listed in the scoring guide or be mathematically equivalent. The figure below shows the percentage spread of the two item types for the ANA tests.

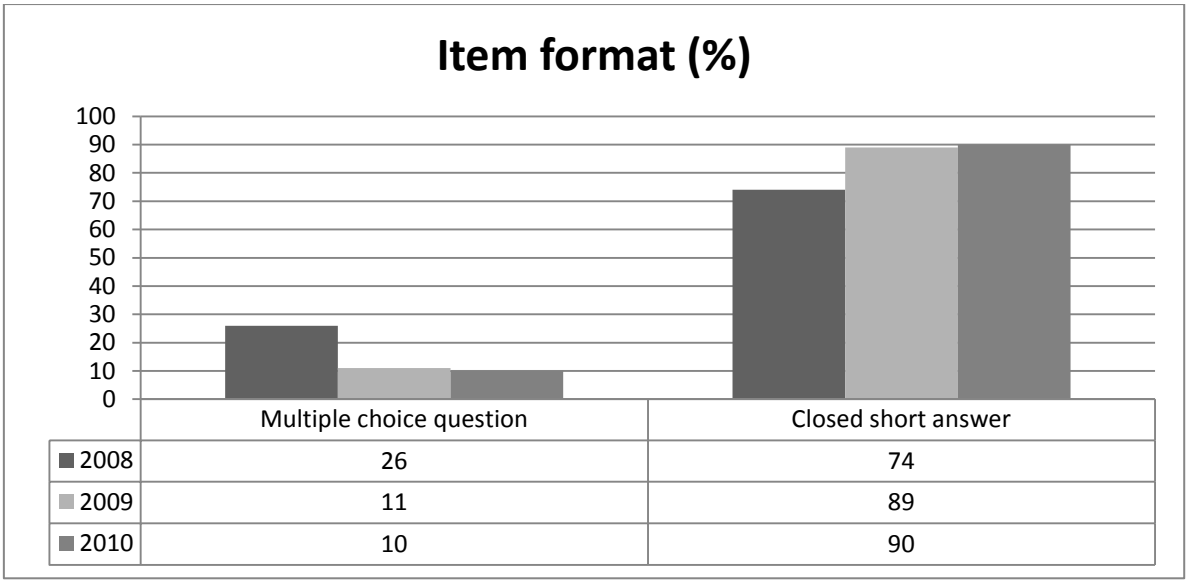


Figure 5.7: IF % spread for the ANA tests: 2008–2010

Figure 5.7 shows the main item format selected by the designers was the closed short answer (CSA) response type where learners had to present their working in the spaces provided or in tables or in graphs. The number of CSA items ranged from 74% in the 2008 ANA to 90% in the 2010 ANA. Across the three assessment years, there was a slight decrease in the percentage of MCQs in the ANA tests. The highest number of MCQ questions featured in the 2008 ANA. In 2009 and 2010, the percentage spread was very similar with almost 90% of questions structured to have a short response. In this way the ANA tests are different from other national assessments (e.g. TIMSS, PIRLS and SACMEQ) where the majority of items are MCQ type. For example, in the 2003 TIMSS study, 77% of the tested items were MCQ, 15% were short response type questions and 8% of items required an extended response type (Anderson and Morgan, 2008). A possible reason for the high percentage emphasis on CSA type questions by the designers could have been the intention to tailor the ANA items format to formats that learners would be familiar in their SBA. Another possible reason is that, according to Withers (2005), multiple choice item writing is a ‘difficult art’ and usually requires several pre-testing stages to generate plausible options with good distractors which can be time consuming and costly.

5.5 Conclusion

In this chapter, focus was drawn on the hierarchical structure of activity and the EMT

analysis. These aspects were elaborated on to give a more informed view on what was found in the assessment artefacts in relation to the design rules advocated for SNA. A clear emphasis was placed on viewing concepts and activities associated with the SA-SNA from an AT perspective.

A departure point for this chapter was consideration of the ANA as a key assessment artefact expected to provide assessment information to a broad community and by so doing, give structure to the SA-SNA activity system desiring a closer alignment between SNA and SBA. In this chapter, it was shown that the goals of SA-SNA can be located with Morgan's assessment discourse frame and provide a context for understanding the anticipated rules for SBA. The structure of this chapter was to take a critical look at the advocated rules evident in policy texts and then critically examine SA-SNA activity system for tensions and overlaps. A significant part of the chapter was to analyse SNA artefacts at item level using specified categories to explain tensions between advocated rules indicated in policy texts with what was found in assessment artefacts.

A critical finding in this chapter was that analysis of assessment artefacts showed significant differences between advocated design rules for the SA-SNA system and the enacted design rules seen in the analysis of the ANA tests from 2008 to 2009. Differences were seen in relation to several design features related to both coverage, and range and scope. Across the three years, the distribution reflected coverage of all five LOs but the closest match to the expected weightings of LO 1 (40%) and LO 2 to LO 5 (15%), occurred in 2008. In the 2009 and 2010 tests, when compared to the expected weightings of RNCS, there was an over-emphasis of LO1 and LO5. For example in the 2009 test there was an over emphasis in LO1 by 9% and in LO 5 by 5% and under-estimates in LO 2 by 8% and LO3 by 4%.

Within the range and scope of questions, there were important observations to note. On DL and CD, the enacted balance of design features show a skew towards the easier end of the frameworks. But lack of NT items tends to push in the opposite direction given the SA evidence. And item format in CSA suggests attunement towards what is more common in SBA context, rather than the trend set in international studies. The following specific tensions showed up on: DL, CD, LD and

IF to the advocated rules:

a) The number of moderately difficult items was the highest across all three tests and higher than the expected 60%. The moderate DL ranged from 73% in the 2010 ANA to 82% in the 2009 ANA.

b) In all three ANA assessments, the CD percentage proportions were different to the expected K (20%), A (60%), N (20%) levels with the most number of items being routine applications. The percentage of routine items in the 2009 ANA (71%) was significantly higher than the other two assessments and the expected 60%. The highest percentage of knowledge items featured in the 2010 ANA but this was 7 points higher than the expected 20%. In the 2008 ANA, there were no items that could be considered non-routine in terms of the CD frame used.

c) Almost no NT computational problems in the strictest sense of the criteria used in the LD frame could be found in the ANA tests of 2008 and 2009. The LD was generally split between high text word problems and those that had low text with a certain instructional verb telling learners what to do. There was a lack of specifications in the advocated frameworks on LD but the design observed in the ANA favoured text based items, which historical SA-SNA evidence suggests South African learners do not do well on.

d) The analysis showed that the main item format selected by the designers was the closed short answer (CSA) response type where learners had to present their working in the spaces provided or in tables or in graphs. The number of CSA items ranged from 74% in the 2008 ANA to 90% in the 2010 ANA which was distinctly different to international assessment norms observed in TIMSS, PIRLS and SACMEQ where the majority of items are MCQ type.

Overall then, while on DL and CD, the enacted balance of design features showed a skew towards the easier end of the frameworks, the lack of NT items tends to push in the opposite direction given the SA evidence. Further, the item format in CSA suggests attunement towards what is more common in SBA context. In the next chapter, a similar analysis frame is applied to the SBA of teachers.

6. School based assessment

6.1 Introduction

In this chapter, the nature of SBA design feature rules is unpacked. The chapter deals with understanding design feature rules advocated in documented texts and then looks critically at SBA artefacts in relation to what is advocated. As in Chapter 5, the position taken is to understand design rules in terms of a hierarchical structure of motives and goals (Leont'ev, 1981) and assessment tool formats that feed into an EMT activity system (Engeström, 1987). To set up the EMT framework for SBA design feature rules, analyses were done to investigate inferred motives and relationships between goals, tools and rules, in each school's activity system. The investigation of the schools' assessment tool formats provided further evidence of motives and goals inlaid into their selected artefacts. The EMT framework allowed me to further examine motives, goals and tools as part of a broader design feature rule activity system with interacting components, where overlaps and tensions exist leading to points of contradictions within what is advocated. Evidence of advocated design rules for SBA was found in national, district, and school policy texts and records. Interview data from teachers provided further evidence on what teachers considered as significant influences on design rules.

The EMT analysis is followed by a mathematical analysis of SBA artefacts in relation to advocated rules. The approach is thus similar to that employed in Chapter 5. The focus on the SBA tools is flagged as central and critical to the analysis of SBA. The analysis is presented using the specific themes and design features foregrounded theoretically in Chapter 2 and exemplified through the additional sub-categories presented discussed in Chapter 5. In this chapter, the analysis is structured according to three themes: the coverage of content, the range and scope of questions and the assessment of common topics. As stated already, the assessment of common topics (which did not feature in Chapter 5) was added to provide in-depth comparative detail on how teachers approached their selections of mathematics on common topics. Within each theme, the relevant design features and sub-categories were used to generate analysis tables which provided a basis for comparisons to be made across the three centres. The approach in this part of the chapter was to consider the SBA portfolio of a specific teacher for the entire academic year. SBA

activities were explored across three schools: a suburban school (S), an inner-city school (IC) and a township school (T). In these schools, selected teachers and their assessment portfolios were the points of reference: Kalay from the suburban school, Fiona from the inner-city school and Mary from the township school (all pseudonyms). Summary tables show the observed overlaps and/or variations within and across assessment artefacts of each school, in relation to what were advocated as design feature rules for SBA.

6.2 The advocated design rules of SBA

The analysis of advocated design rules for SBA was based on documented evidence in texts and tools extracted from policy documents, assessment frameworks, assessment guidelines, the teachers' assessment portfolio files, and interview data relating to assessment design. The description and interpretation of assessment activities in school as activity systems were used to make inferences about motives and goals in each school.

6.2.1 Motives

a) Motives inferred from national policy documents

SBA motives were analysed from: national policy texts such as the RNCS and the FFL assessment framework, and district level texts indicating a mediation of national policy texts and district regulations. The education sector plan was flagged in the previous chapter as another source of SA-SNA motives and also has reference here on the State's advocacy for improving learning outcomes in Mathematics. As indicated in Chapter 2, the RNCS provided the grade-specific scope of work consolidated into 5 mathematics Learning Outcomes (LOs) and the ASs provided teachers with the knowledge, skills, and values they had to cover and that learners were expected to display in a grade. Local influence of these texts and associated motives was evidenced by teachers' direct referencing to LOs and ASs in their assessment planning records.

The assessment guidelines of the RNCS considered the broad purpose of assessment in terms of gathering information to assist teachers in making decisions about the progress of learners. According to the document, assessment informs

teachers' decision making by helping them to establish whether learners are making progress towards the requirements outlined in the ASs (DoE, 2002b). Within this broad purpose, assessment in SBA is seen as an integral part of teaching and learning and should be included in all levels of planning.

The FFL assessment framework provided teachers with quarterly milestones of the LOs and assessment standards (ASs). The purpose was to package the milestones (knowledge and skills) derived from the LOs and ASs in the RNCS into 4 terms for each grade to facilitate assessment planning. The education ministry regarded the FFL framework as a support tool for teachers to monitor progress on LOs and ASs in support of the national goal of improving historically low numeracy levels in primary school grades.

b) Motives inferred from each teacher's Grade 6 assessment documentation

Suburban School (S): Kalay

The document analysis showed that in the suburban school Kalay kept a copy of the RNCS policy document (DoE, 2002b) in her assessment file for ongoing usage and reference. This was her source document to establish what mathematical content should be taught. She also had a reference document to the FFL milestones supplied by the district office with notes on curriculum coverage and the provincial improvement plan (2009–2011). The FFL document included an action plan with provincial targets to increase mathematics results by 30% over a 3-year period.

District texts included several circulars and memos. These listed information that required the school to take note of district intermediate phase management plans, cluster group workshops and meetings. There was some specific information from the district office on assessment workshops and participation in district-based tests. The collective decision of the school management in this school was not to participate in the district based tests.

Kalay reported that the local district office had instituted external cluster moderation of assessments at Grade 6 in all primary schools. With cluster moderation, the aim was for all assessments designed by a teacher to be moderated by a Grade 6

teacher from another school within the same district cluster. The moderation was then affirmed by the cluster leader and a chief moderator. The moderator gave feedback on the teacher's portfolio of assessments and their match to the assessment records of learner performance.

Documentary evidence indicated that the suburban school fell within a cluster of 11 schools and included well-resourced and disadvantaged schools with a mixture of quintile[7] groupings. The cluster met once a term and each meeting had a specific focus and discussions included: checking and verifying all planning documents, how to conduct moderation of assessments, feedback on annual tests and foundations for learning, identifying professional development needs, feedback on competency tests and tracking progress made from the previous term. Meetings focused mainly on monitoring coverage against the policy documents and guidelines.

Inner-city school (IC): Fiona

In Fiona's assessment file, there was evidence of national policy texts such as the RNCS for mathematics as well as the FFL mathematics milestones that she used as reference in her assessment planning. Like Kalay, Fiona had official circulars from the district regarding assessment in her portfolio file. These circulars dealt mainly with information on cluster moderation for Grade 6 assessments and issues relating to the FFL campaign. Reports from the cluster meetings indicated that Fiona's assessments were also moderated by teachers from other schools in the same cluster. There were also indications of several memos from the district calling intermediate phase teachers to centralised meetings to discuss assessment issues. One such district meeting specifically dealt with issues on the administering of the ANA, and Fiona had a manual on the management, administration and quality assurance of the ANA in the portfolio file for reference, although her utilisation of this manual within her SBA was not clear. Again, activities were focused mainly on monitoring coverage against the policy documents and guidelines.

Township school (T): Mary

[7] The quintile status refers to the school's classification according to the poverty index of the surrounding community.

In Mary's portfolio file, there was no evidence of national policy texts such as the RNCS for mathematics or the FFL mathematics milestones, but she indicated that these were available with the school management.

Memos and circulars from the district on assessment were kept in Mary's assessment portfolio file. The district supplied the T-school with common examination papers for all learning areas including mathematics that had to be utilized for quarterly testing of learners. A management plan on the administration of these common examinations was also supplied to schools.

In Mary's case there was specific evidence of district activities that had direct implications for her work. District texts dictating external benchmarking through common district tests and assessment plans were evident in her portfolio file. The district supplied the school with common assessment plans for each term. These assessment plans indicated how work should be covered (work schedule) in the year with content (detailing specified core knowledge and skills) divided into assessment tasks. A start and end period were indicated for each task. In the district assessment plans the activity, content to be covered, LO, AS and mark allocations were specified. The activity options provided to teachers ranged from class work assessments, mental tests, projects/assignments and formal tests. The school had specific dates within which to complete the assessments.

Summary

Of interest were differences between the three schools in terms of motives drawn from national and school-level documentation. For Kalay and Fiona, key influences related to curriculum coverage guided by national policy documents, with school level development of assessments. Moderation of assessments undertaken through district cluster meetings played a secondary role. For Mary, the influence was opposite: the district level documents influenced the specification and schedule of assessment activity, and thus provided the motives for activity. The district level therefore had more direct implications for her work than was the case for Kalay and Fiona – an issue that I take up in the next section.

6.2.2 Goals, tools and rules

a) Goals, tools and rules inferred from national policy documents

The more direct actions of teachers in designing assessment tasks provided the basis for inferring their goals relating to assessment. Goals directing the specific actions of teachers on assessment design were not explicitly stated in the policy texts but can be inferred from details about the form of assessment in the RNCS assessment guidelines for Mathematics (DoE, 2002). The guidelines suggested the importance of assessment tasks that were appropriate for the age and grade of the learners being assessed. Teachers were also advised to approach their assessment tasks with a purpose of equipping learners to work with real-life experiences to participate and contribute to the world in which they live in. In this document, the tools used to assess the various assessment standards were neither prescribed nor fixed, and teachers were asked to make their own professional judgment on the type of assessment that is most appropriate for the purpose of assessment. The need for multiple opportunities for learners to display their knowledge and skills through a variety of formal and informal assessment tasks is noted. According to the guidelines (DoEa, 2002, p.8) teachers' formal assessment tasks should target:

- Full coverage of LOs and ASs in their assessment plan.
- Tasks that over the year reflect the full range of knowledge, skills and values.
- Assessment tools that suitably match the knowledge and skills listed in the LO and AS being assessed.

In the RNCS policy for mathematics (DoE, 2002b), beyond noting content shifts, the following were indicated for teachers as the focus in the Intermediate Phase (Grades 4-6):

- A shift from counting reliably to calculating fluently with all four operations.
- An extension of using different yet equivalent representations to describe problems or relationships.
- A shift from recognition and simple description to classification and more detailed description of shapes and objects.
- A shift from estimation to the use of standardised units of measurements and appropriate instruments for measuring.

- An increased focus on data handling.

Against these goals, teachers had to design assessments to show evidence of developing learners' knowledge, skills, and values and identifying learning gaps in Mathematics. These were investigated in the teachers' assessment planning documents and records (which formed part of their portfolio of documents). The source documents to extract these were school-level texts that showed teachers' assessment plans for the year, the work schedule and the lesson plans. The interviews held with teachers provided further details on how the teachers perceived SBA activities in their school. Evidence on assessment goals was collected in all three schools with summary points referenced to Morgan's (2000) discourse framework.

b) Goals in each school's assessment activity system

Suburban School (S): Kalay

In the suburban school Kalay noted that it was important for teachers to continuously assess and monitor learning over time using a variety of assessment tools. She accessed a range of textbook resources supplied by the school that summarised the year-long content for the five LOs into a single spreadsheet. The type of work covered daily in learners' books was structured according to specific content areas e.g. knowing differences between rectangles and parallelograms. The teacher's assessment plan indicated class assessments and mental tests as forms of assessment. The documentation suggested an intention to test specific maths content in class assessments and a broader range of content in the longer control tests.

Efficient planning on coverage of LOs and ASs was a collective goal in the suburban school. Assessments for the four terms were structured into an assessment year plan. In each term, there were three to four tasks made up of different assessment activities. For example the assessment activities within one task included mental tests and assignments whereas another task included a control test. A control test was a longer test covering different LOs conducted at the end of each term. For each assessment activity Kalay indicated the LO and AS. Her assessment plans formed

part of the composite grade assessment plan in the school which Kalay noted was collectively put together within the school. The assessment year plan was linked to Kalay's work plans (term and weekly). Kalay indicated in the interview that these plans had to be prepared in advance of the commencement of each school term.

K: Our assessments we must know, in *(Kalay's school)*, we must know beforehand what we are teaching beforehand

I: Beforehand, at the beginning of the year?

K: Beginning of the

I + K: term

I: Ok

K: Before the term starts we must know what we are teaching and these are the topics, if it's Decimals, Financial Maths, or whatever, and we need to set our tests well in advance.

The assessment plans had to be approved by the management of the school.

Kalay's work plans noted her use of various texts as resources for sourcing main activities as well as enrichment and remedial exercises within assessments. These were generally photocopied and the teacher put together relevant sections to form questions for learners to complete.

Within the broader assessment activity system, in Kalay's school there was close monitoring of throughput rates and teachers would be reminded not to let their test designs negatively reduce learner's ability to pass examinations and be promoted to the next grade. Kalay re-iterated this caution in the interviews.

K: Sometimes like she *(moderator)* always gave me the suggestion that I shouldn't have my levels too high.

I: Too high?

K: That they will end up failing

I: The performance is low

K: and it's going to be low. So she usually moderates my whole paper and she makes lots of changes.

For each assessment activity, Kalay recorded how many learners were placed into performance ratings using the 4-level ratings scale drawn from the National Protocol

on Assessment document (DoE, 2005d) explained in Chapter 2. This was used to determine the level of learner achievement for internal tests and grade progression. As stated earlier, a pass requirement in terms of this ratings scale loosely translated into learners achieving a level 2 (35%-49%) or higher.

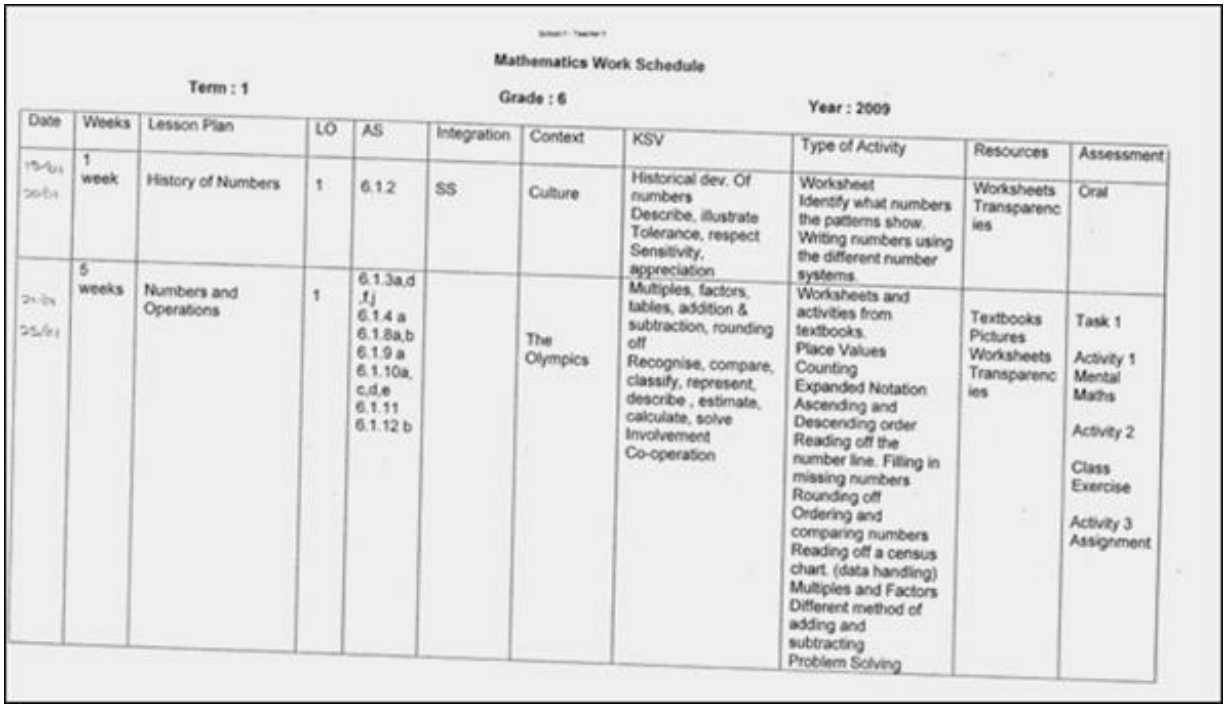
Table 6.1: Pass rate for Mary's Grade 6 classes in term 1 in 2010

Grade	% Passed (level 2 and higher)
6A	86%
6B	58%
6C	62%
6D	67%

In the four Grade 6 classes (6A, 6B, 6C and 6D) that Kalay taught and assessed, the learners were of mixed ability. Performance ratings of learners were fairly high with most learners placed at levels 3 and 4. The Term 1 pass rate trends were similar for the other terms as well. Kalay's file indicated a composite class symbol analysis per term, which summarised comparative class performance.

Inner-city school (IC): Fiona

An important goal for Fiona was efficient assessment planning. Fiona organised and planned work for the year through a learning programme, work schedule and individual lesson plans. In the learning programme there was evidence of topics to be covered, their duration, their match to a LO and AS, integration and context details, resources to be used and assessment. The assessment was structured into oral work, tasks and activities. Curriculum coverage reflected all five LOs with work spread across four terms. The work schedule resembled the learning programme structure with the additional specification of dates included. The work schedule also added further detail on specific skills and knowledge included in classroom activities and the type of test (e.g. mental test, assignment) to be administered. In **Figure 6.1** an extract of a work schedule from the IC-school is shown.



Mathematics Work Schedule

Term : 1 Grade : 6 Year : 2009

Date	Weeks	Lesson Plan	LO	AS	Integration	Context	KSV	Type of Activity	Resources	Assessment
19/01 26/01	1 week	History of Numbers	1	6.1.2	SS	Culture	Historical dev. Of numbers Describe, illustrate Tolerance, respect Sensitivity, appreciation	Worksheet Identify what numbers the patterns show. Writing numbers using the different number systems.	Worksheets Transparencies	Oral
2/02 25/03	5 weeks	Numbers and Operations	1	6.1.3a,d ,f,j 6.1.4 a 6.1.8a,b 6.1.9 a 6.1.10a, c,d,e 6.1.11 6.1.12 b	The Olympics		Multiples, factors, tables, addition & subtraction, rounding off Recognise, compare, classify, represent, describe , estimate, calculate, solve Involvement Co-operation	Worksheets and activities from textbooks. Place Values Counting Expanded Notation Ascending and Descending order Reading off the number line. Filling in missing numbers Rounding off Ordering and comparing numbers Reading off a census chart. (data handling) Multiples and Factors Different method of adding and subtracting Problem Solving	Textbooks Pictures Worksheets Transparencies	Task 1 Activity 1 Mental Maths Activity 2 Class Exercise Activity 3 Assignment

Figure 6.1: IC-school Grade 6 mathematics work schedule

Fiona gave a lot of attention to tracking learner progress. Her portfolio included assessment plans, record sheets of learners' marks, additional class mark sheets, formal assessment tasks and memos, formal cycle tests (that included quarterly tests, half year (June) assessment and year-end (November) assessment), and the ANA. Fiona's assessment plan also provided specific details on the skills and knowledge she targeted in her assessment tasks.

Like Kalay, Fiona designed her own assessments and deliberately sought out a wide variety of resources as reference material to design assessments:

F: I use a lot of textbooks, the resources I use, I don't just use one book, I use a variety of books, get my information off from the internet – local supplements that we get we use activities and things from them.

Fiona was particularly concerned with standards that learners were expected to reach in her assessments but had no formal SBA guidelines from the policy texts and the school to assist her establishing a spread of difficulty levels and cognitive demand in her designs. While a framework for coverage was evident, as with Kalay, there was no evidence of a pre-defined framework that guided the structure of the

tests and Fiona relied on her own subjective experience in setting papers that had a balance of difficulty and cognitive spread.

Township school (T): Mary

The key goal observed from the documented evidence in the township school was for Mary to align with district plans and to utilise tools supplied by the district. The district goals were directed towards supply and provision of pre-packaged curriculum and assessment schedules and tools to schools that were historically under-resourced (i.e. with a low poverty index quintile ranking) that were described as aligned with the expectations of the curriculum. The district goals were systemic. The primary goal was to mediate coverage of the RNCS through pre-set planning documents and tests. Hence, the package supplied by the district, included common work schedules and term-end assessments. A second stated goal was to monitor learner progress in light of the strategies and interventions of the district and these were contributing towards the national goal of improving numeracy levels. This led to annual school target setting over a certain time period as indicated in the FFL campaign.

The goals of the district were articulated through district circulars and memos, and workshops held with teachers. The district also supplied the school with common assessment plans for each term. These assessment plans indicated how work should be covered (work schedule) in the year with content (and its specified core knowledge and skills) divided into assessment tasks. A start and end period were indicated for each task. However, there was no clear documentation or framework supplied to teachers on the design of the common assessments, limiting the possible "backwash" effect on teachers' assessment design in the township school.

Evidence from Mary's portfolio file revealed that although she did create some of her own assessments, there was no evidence of deliberate intention to align these to the district texts. The district plan was also not formally used to indicate completed tasks. Mary indicated in the interviews that she was guided by the district schedule but did not always stick rigidly to it.

I: OK. So…this plan of assessments where the district has stipulated the type

of assessment, classwork, this is the date, this is the month.

M: We use the assessment, ja

I: You use this as a guide

M: Yes

I: to complete your assessment?

M: Yes

Later, in the chapter, an analysis compares district common tests and Mary's own assessment design conducted in the township school.

In Mary's school, there was evidence of more bureaucratic compliance to instructions received from the district. Teachers completed formal term schedules, which indicated learners' scores in mathematics. These schedules were signed off by the school principal and handed over to the local district for each term. District officials also conducted on-site visits to the school to check teacher files and monitor compliance towards the district work schedules. Like Kalay and Fiona, Mary made use of individual class sheets to record learners' performance on individual activities. For each term, there were two tasks recorded that were broken up into several subcomponents or activities. These subcomponents comprised: mental tests, homework, class work, projects/assignments and tests, an indication of further compliance to the RNCS assessment guidelines.

Summary

Across the three schools, the evidence from the activity systems showed curriculum coverage compliance driven goals. In the suburban and inner-city schools, the inferred goals were similar with a clear emphasis towards monitoring and assessing learning over time using their own assessment, and overt emphasis on coverage of LOs and ASs listed in the national curriculum statements. Both of these were listed as expectations in the assessment guidelines. In relation to Morgan's assessment discourse framework, these goals leaned towards the "curriculum implementation" strand. There was limited influence by the district on these goals. At school level for Kalay and Fiona, there was the added pressure of throughput and high pass rates for teachers to consider. In these schools, evidence of goals came through activities related to assessment planning, compliance to policy texts, wide use of resources to

formulate questions and tracking throughput at class level. However, there was more sense of goals driven by collective school-level assessment activity in terms of expectations and moderation for Kalay, in contrast to school level expectations but more individual working on assessment for Fiona.

In the township school, the inferred goals from the activity system were about compliance with district-level insertion of goals, tools and rules. The principal focus of the district was on promoting systemic goals. In this regard, the school participated in all the common tests supplied by the district. There was less emphasis in the documentation on teacher development. In relation to Morgan's discourse framework, and information from interviews, while Mary's school goals technically fell into the "curriculum implementation" strand, these were both dictated and circumscribed by district documentation and its limitations. In later analyses in this chapter, evidence is provided to show that there were disjunctures in Mary's assessment tools and the district goals as mediated through common assessment plans and common test papers.

The lack of evidence of a test design framework is worth highlighting at this point. Across the schools, the activity systems showed that the teachers could not extract information on assessment frameworks and design rules from the policy texts either at national or at district levels, and had to rely on their own subjective experiences in setting "cognitively balanced" test papers. For Kalay and Fiona, there was some evidence of attention to cognitive balance through school level expectations related to throughput.

I go on now to look at the outline of assessment tool formats seen in each teacher's assessment portfolio. The analysis of goals and tool formats, combined with documentary and interview-based analysis of rules within each school's assessment system are then combined in the setting up of an EMT model for each school's activity system.

6.3 SBA tool formats

This study investigated how motives and goals influenced the design feature 'rules' of teachers' SBA. As indicated in the introduction, a critical part of the investigation

into the SBA of teachers, were the actual assessment tools in teachers' portfolio files, interpreted as manifestations of motives and goals inlaid into their selected artefacts.

My initial focus is on tool formats given that the RNCS assessment guidelines prescribed that teachers could make use of the following range of assessment tools: tests (informal and formal), homework, assignments, projects, investigations, investigations and examinations (DoE, 2002a). For each teacher, the assessment tasks with categories as they appeared in portfolio files are indicated. Below each assessment the allocation of marks is indicated. This gave a sense of the relative weighting of the assessment tool in relation to the others listed. All assessment tasks for a single academic year, spread across four terms, were covered in this analysis. For ease of reference, the specific assessment tools listed in **Table 6.2** were coded according to the relevant teacher and assessment tool. For example Kalay's mental test was coded as K1, class tests as K2, assignments as K3 and so on. In the table, the exact labeling used by the teachers is maintained but some categories meant different things for different teachers (e.g. assignments and projects). Brief descriptions of the tools designed by the teachers are provided.

Suburban School (S): Kalay

Kalay made use of various types of assessment tasks. These included mental tests, class tests, class work, assignments, projects, investigations and control tests with this range summarised in **Table 6.2**.

Table 6.2: Assessment tasks for Kalay

Term1			Term 2		Term 3				Term 4		
Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9	Task 10	Task 11	End of year
Mental test	Mental test	Cont.[1] test	Class work	Speed test	Proj.[3]	Assig.[1]	Cont. test	Class work	Invest.[4]	Class test	Cont. test
20	10	80	25	10	15	20	30	40	20	45	45
Mental test	Class test		Assig.[2]	Speed test	Mental test	Mental test		Assig.[2]	Invest.[4]		
10	40		20	10	15	10		25	30		
Mental test	Class work		Assig.[2]	Control test	Class work						
10	20		20	40	30						

Key: 1. Control Test 2. Assignment 3. Projects 4. Investigation

K1) mental tests and speed tests

Mental and speed tests were short tasks designed around specific ASs. Several mental tests were indicated across the four terms. In a single term, as many as four mental tests could be written by learners. The mental and speed tests were structured around a specific computational skill e.g. addition of two whole numbers.

K2) Class work

Class work assessments were short activity based tools. They were based on a specific skill or topic and they took place during class work time periods. The structure of these tests were similar to class tests but drew on a smaller range of work.

K3) class tests

Class tests were designed by Kalay to assess learner competencies through a more formal approach. These were scheduled tests and were based on more than one AS and had a longer duration than the mental tests. Kalay designed class tests with section headings e.g. Number operations and relationships to indicate the broad content area being assessed.

K4) Control tests

Control tests were formal assessments given in each term. Kalay assessed all the LOs covered in that term. There was a selected examiner (e.g. Kalay) and moderator to check the quality and standard.

K5) Assignments/Projects

Projects and assignments were designed to be completed over an extended period outside of the class activities. Included in the design was assessment of learners' ability to select an appropriate strategy (method, tool, and technique) for the LO and AS being assessed. Assignments were assessed on quality of the mathematics content, presentation, originality, design and bibliography.

K6) Investigations

Kalay also formally assessed learners' mathematics capabilities through

investigations. This included applications on more non-routine and problem solving type questions (e.g. number pyramids, magic numbers, magic geometry) focusing on multiple ASs in a LO.

Overall, from the SBA tools Kalay designed, the control test had the highest weighting in a term compared to other categories. The control test was similar in format to an examination.

Inner-city school (IC): Fiona

Fiona's assessment plan provided details on her assessment tasks. A single task was comprised of subcomponents of various formal activities counting as part of the assessment task. These subcomponents ranged from a minimum of 3 to a maximum of 6 activities. A single assessment task could include a mix of speed and accuracy tests (mental), assignments and formal class tests. On the assessment plan the nature of the skill (e.g. multiplication and division) being assessed and the associated assessment activity (e.g. class test: long multiplication) were indicated. Fiona had compiled an assessment plan for each of the four terms of the year with two tasks indicated for each term. In addition to the two tasks per term, all learners wrote a cycle test at the end of each term. Fiona made use of a variety of assessment tools in her SBA. **Table 6.3** indicates the various assessments she conducted across the year with short descriptions below

Table 6.3: Assessment tasks for Fiona

Term1		Term 2		Term 3		Term 4		
Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	End of year
Speed and accuracy	Class work	Class work	Class work	Class work	Class work	Class work	Class work	Exam[4]
20	10	10	10	10	15	15	23	40
Class test	Class work	Class work	Class work	Class test	Class work	Class work	Class work	
30	5	20	15	20	15	15	25	
Class work	Cycle test	Class work	Class test	Class test	Assign.[1]	Class work	Class test	
10	25	20	38	20	20	10	25	
		Assign.[1]	Assign.[1]		Speed and accuracy	Assign.[1]		
		30	22		30	20		

		Class test	Exam[4]		Cycle test	Class work		
		20	25		20	10		
						Class work		
						15		

Key: 1. Assignment Homework 2. Investigation 3. Project 4. Examination

F1) Speed and accuracy tests (mental)

Fiona's speed and accuracy tests were structured in the same way as Kalay's mental tests with short tasks testing number bonds and operations. They usually comprised 20 questions and were presented on an overhead projector with learner responses in their workbooks.

F2) Class work

Fiona's most commonly used assessment tool was class work. As with Kalay, these were short specific assessments based on a specific skill (e.g. time conversions or minutes to hours). Fiona structured these as activity based assessments that were conducted as soon as a specific skill was complete. These shorter class work assessments provided the basis for her slightly longer class tests.

F3) Class tests

Fiona structured her class tests on individual topics (e.g. adding and subtracting using the column method). Class tests were short but were based on a cluster of skills within a single AS or different ASs.

F4) Cycle tests

At the end of term 1 and 3 Fiona conducted a formal cycle test with learners. This covered a broad range of LOs and ASs covered in the term, thus resembling a formal examination. The cycle test included mark allocations per question and space for answers to be written. Learners were given written instructions regarding their answering of questions and questions appeared in different formats (diagrams, tables, closed response).

F5) Assignments

Fiona's assignments were structured similarly to class tests in terms of work

coverage and were based on a specific skill. Assignments were printed for learners and activities were separated into sections. A section on “problem solving” was always included, with more tasks involving multiple ASs.

F6) June/November Examination

The June/November assessments were similar in structure to the end of term cycle tests with the additional inclusion of an examiner and moderator. Coverage was based on what had been done in the second and fourth terms respectively. The format of the assessment was divided into sections with content areas (topics) specified.

Overall, although Fiona made use of a variety of assessment tools, the weightings across them did not differ as much as Kalay. For example, Fiona’s class test, cycle test and the examination had similar weightings, giving short term assessment a larger weighting overall in comparison with the S-school.

Township school (T): Mary

In Mary’s assessment portfolio, there were two tasks recorded in each term that were broken up into several subcomponents or activities. These subcomponents comprised: mental tests, homework, class work, projects/assignments and tests. **Table 6.4** indicates the frequency and range of Mary’s assessments across the four terms with short descriptions below.

Table 6.4: Assessment tasks for Mary

	Teacher Z							
Term1			Term 2		Term 3		Term 4	
Task 1	Task 2	Task *	Task 3	Task *	Task 4	Task *	Task 5	Task *
Mental test	Mental test	District test	Class work	District test	Class test	District test	Class test	District test
20	20	50	30	50	40	50	50	50
Mental test	Class work		Mental test		Mental test			
05	05		20		16			
Homew.[5]	Homew.[5]		Class Test					
08	08		40					
Class work	Project							
05	22							

Assign.[1]	District test							
20	50							
Class Test								
50								

Key: 1. Assignment Homework 2. Investigation 3. Project
4. School-based examination 5. Homework
* District common paper but not specified in Mary's assessment plan

M1) Mental tests

Mary designed mental tests on specific topics (e.g. polygons). These mental tests were prescribed assessments drawn from the assessment term plan supplied by the district for each term where content and frequency were pre-defined, but item design was left to Mary.

M2) Homework

Homework was given to learners on a regular basis and was usually regarded by Mary as an extension of class work activity.

M3) Class work

Class work was a prescribed assessment and had the same weighting as the mental tests. Class work was assessed on a daily basis and was based on a specific skill (e.g. expanded notation of numbers) for learners to complete.

M4) Assignments and projects

In each term learners were given either an assignment or project to complete. Unlike Kalay, assignments and projects (like homework) for Mary were seen mainly as an extension of class work, so while named differently, tasks were similar to M2 tasks. For assignments and projects the district specified the intended content to be assessed (e.g. financial mathematics and problem solving) but there was no direct correlation to these topics in Mary's assignments and projects. Assessment tasks indicated that assignments and projects were selected from textbook resources and given as photocopied worksheets to learners. Explicit criteria for assessment were not indicated. The content included in these assessments was often based on more investigative work (e.g. mapping activities) similar to Kalay and there were indications of learners working in groups to complete the tasks. Assignments and

projects were less frequently given than class work and homework activities but they contributed a high weighting (15%–20%) in the term mark of learners.

M5) District tests

District tests were set as common papers at the end of a term. They had a slightly higher weighting than the formal class tests given to learners. The district test was a summative assessment of tasks completed in that term. For this reason, work was aligned to the common work schedule supplied to schools. The time allocation was one hour and the questions were based on specific topics which were listed in the test (e.g. history of numbers). For each question the associated AS was listed.

M6) Class tests

In addition to the common district test, Mary independently designed a few formal summative class tests for learners that fell outside the district requirements. These tests had the same weighting as the district test but did not cover as many questions as the district assessments. The tests were also an hour long and the ASs was listed on the cover page. Some tests were handwritten with photocopies given to learners.

M7) ANA

There was also evidence in Mary's portfolio that her learners wrote an ANA paper as a formal task in the fourth term. The school management team had decided to include the ANA as a formal task in Grades 1–6 where an ANA paper from that year was available. This was done in 2009 but not in 2010 because the ANA paper was not available in the fourth term. Therefore the ANA was excluded from Mary's SBA analysis.

Overall, while Mary's class tests and the district tests had higher weightings than other categories across the year, there were fewer assessments in later terms of the year and a more limited range of tools compared to Kalay and Fiona.

Summary

The documented evidence from the assessment plans indicated that all three teachers complied with the minimum policy requirement of having two assessment tasks per term. Kalay and Fiona made use of various types of assessment tasks

which in turn were divided into sub-components. Fiona had the highest number of sub-components in an assessment task. These subcomponents ranged from 3 to 6 different assessment activities, for example, speed and accuracy tests (mental), assignments and formal class tests. The frequency and range of Mary's sub-components were less extensive than for Kalay and Fiona but included district tests as a summative assessment of tasks completed in that term. District tests were not indicated in Kalay's and Fiona's assessment plans. Kalay did give more weighting to longer-term assessments (e.g. control assessments – Task 3) than other forms of assessment.

6.4 Assessment activity within an EMT framework

I use the analyses above to describe components of activity systems in each of the three schools in terms of Engeström's (1987) EMT model, In Chapter 3 this model was explained and considered a useful framework to understand assessment activity systems. This allows me to consider key tensions and contradictions within the three schools' assessment activities, which, in turn, allows for an understanding of the broader assessment activity context within which mathematical assessment tasks were designed or used.

Kalay and the suburban school (S)

a) The subject

Kalay had 13 years teaching experience. In 2010, she had her third year of teaching Grade 6 Mathematics. She had taught Mathematics for the past 10 years of her 13 years teaching – mainly in the senior phase. In the interviews she commented that she was "comfortable" teaching Mathematics to senior phase learners and was "actually enjoying it" at the school. She was qualified to teach Mathematics in the senior phase (Grades 7-9). She was involved in school committees – one of which was the school assessment team, where she had served for two years.

b) Object

The object of SBA for Kalay as gleaned from the interview data was to assess work covered according to the requirements of the policy statements on curriculum, namely the RNCS and the FFL. The portfolio contained school policies, RNCS

assessment standards and learning outcomes and reference to the FFL document. These documents were used as a reference for the expectations of the curriculum in assessing learners in Grade 6 mathematics. Kalay indicated in the interviews that the FFL milestones curriculum framework was used as a guide but she found it difficult to use.

I: *OK, so you use it as a reference document?*

K: *It's not very user friendly in that it's very, it's a bit vague.*

Kalay complained about the packaging of content in the FFL documents that made it difficult to follow in practice.

K: *Because here you touch on fractions every term, a little of each and then I find like they go straight into decimals without teaching fractions first, so you can't really follow it very strictly …Cos certain things it just doesn't work, like you cannot teach decimal fractions before actually teaching fractions. And they touch, you know, bits, with bits each term…*

Kalay based her assessments according to a work schedule that she developed using the FFL documents and district guides to ensure she complied with policy requirements. She cross-checked the coverage of content in the work schedule against some textbooks.

K: *So when I look at all the Grade 6 textbooks then I say, 'OK, this is what's expected and…' Textbooks help a helluva lot.*

Kalay indicated that she had received some helpful documents from the district that she also used as a reference, documents that could assist in standardizing practice in the district:

K: *But they do help you in terms of, like the mark allocations and stuff. I think like I find that you need more guidance in terms of, you know, how much percentage, how much should be tested on this and that, so we're all like, you know, everyone's standardized.*

c) Assessment tools

Kalay used a range of assessment types that included mental tests (which she sometimes labeled speed tests), class tests, control tests, projects and assignments

and investigations. Varied content application was observed with varied difficulty, and cognitive levels were more explicit in the formal examinations and class tests. The tools showed extended variation in example space within topics. This is explained in greater detail in the analysis of common topics.

d) Rules

In terms of assessment design, subject specific meetings happened quite regularly at the school to discuss and address matters relating to the school, district and national policy requirements. The assessment culture of the school was highly regulated and assessment tasks had to conform to standard rules of the school relating to broad coverage in assessment of national curriculum. For example, all formal tests had to be moderated by the head of the department (HoD) before being administered to learners. Often, the principal performed a high level check as well.

On coverage, documentary evidence showed that Kalay grouped assessment standards into mathematical topics (e.g. natural and whole numbers, common fractions, 2-D shapes and 3-D objects, and data). Some links were also made to the textbook spreadsheet which also seemed to be a resource that guided planning.

Kalay also expressed school expectation-related norms of high pass levels and the need to work across the learner attainment range in support of learner throughput.

e) Community

There was an established sense of community at the school. Within the mathematics departments, Kalay was at ease to consult senior and peer teachers for advice on assessment tasks and appropriate resources to use. She usually consulted her peers before the tests were handed over to the HoD for approval. The subject meetings provided a forum for Kalay to receive feedback on assessment tasks. Kalay received formal feedback from the HoD on her assessment tasks and feedback documents were kept in her portfolio file. There was also a school-level rotation policy with each staff member given a chance to serve on various committees. Some of the activities of the school assessment team included: drawing up a common exam timetable and assessment year plans.

The district facilitated the hosting of assessment workshops where school clusters congregated. Within these clusters, schools were twinned to check and feedback on assessment tasks. Kalay described the feedback as mainly structured on compliance rather than on qualitative feedback on the assessment tasks:

I: Ok. So … But other than that, the district or the clusters haven't really met to discuss your assessments?

K: No

I: Ok. So it looks like most of, most of what you are doing – assessment, your tests, your assignments – is it, is it a product of your engagement with other Maths teachers at the school?

K: Yes …and our HoD…usually calls for our assessment files, you know, well in advance, and he moderates it and one feels that, you know certain things need to be changed, so that's …

Parents received regular feedback on the SBA administered with learners and on a termly basis, received a school report on progress.

f) Division of labour

In the S-school, individual teachers were responsible for the design of their assessment tasks. The frequency of formal testing was standardised according to a school assessment plan and formal examination timetables for mid-year and end-of-year examinations. Senior and peer teachers conducted moderation on assessment tasks within a grade or phase, and the HoD for mathematics provided formal approval for the administration of tests. School-based subject meetings included discussions, feedback and editing of specific tasks based on balancing cognitive demand with difficulty levels.

g) Summary activity system table for Kalay in the S-school

Table 6.5 below summarises the EMT analysis for Kalay and the S-school**.**

Table 6.5: Kalay's SBA Activity system

EMT Model	Assessment practice in the S-school
Subject	• Suitably qualified and experienced. • Participates on assessment committees, and able to consult and utilise advice offered by senior teachers and peers.
Object	• Clear motive on completing curriculum coverage of all major topics and

	skills in mathematics. • Assessments had a clear reference to the LOs and ASs of the RNCS but less so to the FFL milestones. • Used textbooks and the feedback from moderators as a guide to plan and develop tasks. • Was interested in designing questions for learners with varied ability levels.
Tool	• There are six types of assessment tools used by Kalay. The most frequent tool used is the mental test which she also labels speed tests. • Either a major assignment or control test is conducted with learners at the end of the term. Control tests have a significantly higher weighting than the other tools used. • Tools (e.g. class tests) are designed mainly through the use of textbooks, past year papers and own knowledge.
Rules	• Assessment planning and designing tests is done in accordance to accepted school policies • There is compliance to the advocated 'rules' relating to expectations of broad coverage in terms of ASs and LOs from the RNCS. • There is compliance with school expectations of high pass levels, and the need to work across learner attainment range in support of learner throughput.
Community	• There is a clearly defined hierarchical structure of management and assessment oversight which is determined by senior management and accepted by teachers of the school. • There is a highly involved school community structure of school managers, moderators, and senior and peer teachers that give support to the assessments designed by Kalay. • District support is available but not considered essential.
Division of Labor	• While each teacher is responsible for the design of their internal class assessments, a structured moderation approach has to be followed. Kalay receives formal feedback from the HoD on her assessment tasks. • The power of assessment design and item selection lies within the school, with national level artifacts and textbooks used to check broad coverage. This creates inherent tensions for Kalay between wanting to push mathematical range and upsetting the passing proportion, which her school is keen on protecting. • The district is not an active 'player' in influencing Kalay's assessment designs.

h) Contradictions for the S-school

Primary contradiction: Object

A key observation is that there are multiple objects in the activity system. Kalay's object is expressed in terms of a match with curriculum expectations on exposing learners to a full range of knowledge and skills, but the school has demands for high

pass rates and throughput in their system.

Secondary contradictions: Subject ⟷ Community

Following on from a divergence in objects in the S-school, Kalay's main assessment objective was to expose learners with questions and tasks that match the ASs listed in the curriculum statements. However, after moderation, her test design was pitched more according to the level of the learners' knowledge in mathematics than to the curriculum requirements. She noted that her initial assessment designs and expectations were sometimes in conflict with those of the moderators.

Secondary contradictions: Object ⟷ Rules ⟷ Tools

This contradiction arises out of Kalay's expectation of assessing learners to the curriculum standard but at the same time being conscious of the psychological needs of the learners and their ability to answer questions of varying difficulty. This affects her design of tools. The moderation structures and performance of learners form key components of the rules that govern the final design of the assessment tools.

Feedback provided from the senior teacher and the HoD indicated that there were other levels of engagement with the assessment tools designed by the teacher. A clear hierarchical structure of moderation was in place to check appropriateness of tasks and questions in relation to the perceived mathematical knowledge of the learners and this sometimes was in conflict with her initial design resulting in some of the more difficult questions being made easier. The contradiction is subtle, but can be regarded as a disturbance that causes Kalay to re-evaluate her design of tools in ways that align with the rules of the school community.

Fiona and the inner-city school (IC)

a) The Subject

Fiona had been teaching for 10 years. She was, at time of the study, the head of department (HoD) of Mathematics at the IC school. She had 9 years' experience of teaching Mathematics and other subjects in the intermediate (Grades 4 to 6) and senior phases (Grade 7). The bulk of her teaching workload comprised of teaching

Mathematics in Grade 6 and Grade 7. Her involvement in school structures included being part of the management of the school, the school's assessment team (SAT), the school based support team (SBST) and the school development team. As a HoD, she was also responsible for coordinating all the subject specific meetings that fell within her supervision, one of which was Mathematics.

b) Object

The key object in Fiona's SBA was to ensure that different ability levels were being accommodated in her assessment tasks. As noted earlier, Fiona expressed tensions in achieving a balance between different assessment discourses (Morgan, 2000) of curriculum implementation and psychological needs of learners. To accommodate her goal of catering for learners with different ability levels, she made use of multiple assessments and sometimes learners could repeat an assessment to improve their marks. Like Kalay, she was influenced by two considerations: the 'type/s' of learner/s (referring to mathematical aptitude) that were present in the grade and the sections of work covered in the term.

F: So depending on the type of learner we get, my assessments are based on that. I look at what my sections are for the term that I needed to cover and learners are assessed accordingly.

In her assessment portfolio file, overt references indicated that she had made use of several policy documents and guidelines in her planning for assessments for the year. Compliance to national policy texts was evident in specific references to the RNCS and the assessment framework of the FFL. During the interviews, Fiona indicated that she had received training on implementing the RNCS. She had also read up on the FFL and gained some insights on it during cluster meetings from teachers from the other schools who had been trying to implement it. In addition to the policy documents, Fiona made extensive use of textbooks to establish the appropriate standard of setting assessment tasks. Although some of her textbooks were directly aligned to the RNCS, she said she preferred to use some of the "older" textbooks for more detailed explanations on certain topics in Mathematics, but used the "newer ones" to ensure alignment with curriculum requirements. The use of textbooks is further explained under Rules.

c) Tools

Fiona, like Kalay, used a variety of assessment tools in her SBA that complied with the national requirements. Her assessment tools included speed and accuracy tests (a form of mental tests), class tests, cycle tests, assignments, and examinations (mid-year and end of year). The assessments appearing in the learners' activity books correlated to the ones featuring in her assessment plans. A key observation was a much heavier presence of class work in comparison to Kalay and Mary, providing learners multiple chances to achieve an AS. Fiona made use of the class work tool 18 times out of the 34 assessments she conducted, re-enforcing the psychological discourse view of providing learners with multiple opportunities in assessment. In the interviews, Fiona indicated that the high frequency of class work assessments provided a handle on where specific problem areas existed so that there could be quicker remediation. This meant a greater weighting towards single AS-focused assessment in her SBA.

d) Rules

School norms dictated that every teacher had to have a work schedule, learning programme and an assessment plan for the year. Fiona's assessment portfolio file displayed these three levels of planning in an integrated way with work schedules and learning programmes making reference to assessments. For example, the topic on numbers and operations in the work schedule and learning programmes included an indication of the type of assessment she would be using, what skills would be assessed, the LO and the AS to be covered, and tentative dates as to when those topics would be covered. Fiona generated topics based on the AS and then used various textbooks to select the sequence of content to be taught and assessed with learners.

As indicated earlier, there was no evidence of a pre-defined framework that guided the structure of the tests. While Fiona noted the use of the Grade 6 work schedule that detailed what topics needed to be assessed and when, this schedule gave no indicators on a pre-determined framework for setting tests.

I: Is there something that tells you, 'Ok, I'm going to end up with a test where

40% must be basic knowledge questions, 20 questions are going to be multiple choice, 10% of my questions are going to be problem solving.' You don't work with those pre-defined categories, like a framework?

F: No I do. I do have… I do write down your …, what I'm going to test, whether it's…

I: Where do you write this down?

F: Oh no, when I'm preparing I just write it down on a page where I put my…

I: On a page?

F: Ja, on a page and I write down my topics and then I write down how am I going to address these topics. Is it going to be multiple choice, is it going to be a true or false?

The design feature rules used in the actual test construct were drawn from a variety of textbooks with some reference made to past years' papers and her prior experience. While Kalay described within-assessment differentiation tailored to classes, Fiona felt that the differentiation within assessment that was being incorporated was not working to meet the needs of the various ability groups in her class.

e) Community

Fiona was more of a "lone player" in constructing her assessment tasks at the school level. Formal assessment tasks had to be moderated by a member of management but the checks were more in terms of compliance than on the specific mathematics topics being assessed or the selected items.

F: Once it's been moderated and approved by the head of department, and then the deputy, then it's then given back to me … get it typed out, copied … the children write the thing.

In her assessment files, community evidence included district invites, notices, cluster meetings and workshops. Fiona attended several cluster meetings where district officials supplied further information on implementing the FFL framework and how moderation of assessments should be done. At some of these workshops, schools were teamed up with other schools to check each other's' assessment files and

examples of learners' work. A checklist given by the district officials had to be completed. According to Fiona, the emphasis was on compliance on number of assessment tasks set by teachers rather than on the nature and quality of questions contained within an assessment. Over the data collection period of 12 months, Fiona indicated that not much had changed in this regard.

F: Not much changes has taken place since then. We're still left to do things on our own.

District feedback was minimal but included confirmations that their assessments were aligned to policy requirements.

I: So there were some visits by the district at your school. And how was the feedback, in particular around the grade 6 maths?

F: The maths facilitator was quite happy with what he had seen and he commented that there's quite a lot of work being done with regards to the grade 6s and that we were keeping in line with what the district required us to do and what policy had stated.

Overall the focus for Fiona was on attendance at district level while at school level she operated as a "lone-player".

f) Division of labour

As long as Fiona complied to each of the levels of planning indicated in the RNCS teacher guidelines (DoE, 2003a), she was relatively "free" to set and design assessments on her own. Being a HoD, her tests were subjected to moderation by the principal or deputy principal but the feedback she received was relatively informal, mostly orally provided and lacking specifics on mathematics content.

As with the S-school, the internal policies of the IC-school determined the workloads and job allocations of individual teachers. Individual teachers had to complete a set number of assessment tasks for mathematics. The frequency of formal testing took place according to the term plan of cycle tests and/or examinations. There was minimal evidence of engagement of teachers with senior or peer teachers. Tests had to be moderated by the HoD for mathematics. Minutes of meetings contained in Fiona's file indicated that regular subject meetings were held to discuss school

requirements for lesson planning, assessments and feedback to parents.

g) Summary activity system table for Fiona

Table 6.6 below summarises the EMT analysis for Fiona in the IC-school.

Table 6.6: Fiona's SBA activity system

EMT Model	Assessment practice in the IC-school
Subject	• Suitably qualified and experienced in teaching Mathematics. • She has a preference to work on her own within the policies of the school.
Object	• Multiple assessment discourses present: curriculum implementation and psychological strands. • Fiona is challenged in her assessment design by finding the correct balance of curriculum coverage and offering expanded opportunities to learners of different abilities.
Tool	• Six different types of tools were used with clear reference made to the LO and AS from the RNCS in the assessments. • Fiona's high prevalence of 'class work' assessments within the overall weighting of her assessment was far greater than such occurrences for Kalay and Mary.
Rules	• Works within the confines of the RNCS but makes also makes use of the textbooks to establish the rules on assessment. • No pre-defined framework for setting of tests.
Community	• To a large extent works individually (as a 'lone player') on tasks but receives informal feedback from the school management. • District feedback and community support is available but not specific to mathematics and is not highly valued.
Division of Labor	• While there are regular subject meetings, assessment planning is the responsibility of each teacher. • Tests have to be moderated by the HoD.

h) Contradictions for the IC-school

Primary contradiction: Object

As indicated earlier, Fiona was largely a "lone player" in constructing her assessment tasks at the school level. There is a contradiction in terms of expectation of common assessment based on curriculum coverage, and her more psychological and individual discourse. As long as Fiona presented her assessments in line with the general school requirements, she was relatively "free" and on her own to set and design the actual questions and tasks on mathematics content.

Secondary contradiction: Community $\longleftrightarrow$ Division of labour

Fiona has limited channels to consult on her assessment designs even though she felt challenged (like Kalay) in her assessment designs on finding the correct balance of assessing to the expected standard of the curriculum while offering expanded opportunities to learners of different abilities. At the IC-school, the community structures (e.g. the district office and school moderators) were in place but ineffective in practice. There was quite limited substantive professional support at the school regarding assessment design and content selection. While Fiona was able to design her tasks on her own, development and support potential are neither emphasised nor realised.

Mary and the township school (T)

a) The Subject

Mary, in the T-school, had more than 7 years teaching experience. She had started her teaching career in the Adult Basic Education and Training (ABET) sector and also had some experience teaching in a private school. In the ABET sector she taught Mathematics level 4, the equivalent of Grade 10 in the public schools. At the private school she had taught Mathematical Literacy and Natural Science at high school level. At the time of the study, she had been at T-school for two years teaching Mathematics at Grade 6 level. Since she was relatively new to the school, she said she was not yet fully involved in school committees and seemed uncertain about their structure and existence.

b) Object

Neither Mary's documents nor her interviews indicated explicit assessment-related objects. In more general terms Mary's object was to cover the required amount of work as stipulated in the district assessment work schedule because it was compulsory for her learners to participate in the common examinations. The district had tried to regulate assessment practices among schools through the distribution of common work schedules (pace setters).

I: *Do you design those tests on your own or are they given to you or do you all design?*

M: *To the cluster we design the assessment plan to the cluster meeting.*

I: OK. So in that… So are you saying then these are tests that are designed by the cluster and then shared amongst you in the schools?

M: Yes, yes. And the homework also – all the activities for term 1.

The work schedule listed the ASs and topics to be covered in a term. This suggested that the district goal was to further mediate the RNCS and the FFL so that adequate coverage for preparation for standardised assessments such as district common papers would be facilitated. The FFL documents were used by Mary mainly as reference documents for the district supplied work schedule.

Textbooks were the preferred source for Mary to understand the required rules that she needed to assess on specific topics.

I: What I want to try and find out is … are you using a textbook to help you determine the assessments that you are going to do with your learners? Does that textbook help you in that regard?

M: Yes

I: Do you take … do you take assessments out of the textbook and give to your learners?

M: No, I'm using others from the textbook. Others I'm using mine.

I: You're … When you say yours, you mean your knowledge?

M: My knowledge.

Mary indicated that she relied a lot on text books to generate tests but also relied on her own knowledge of mathematics to design some tasks. She referred to National texts such as the RNCS and the FFL as reference material but textbooks were the main sources for assessment items that were going to be used.

c) Tools

The tools Mary used in her assessments were a combination of the district assessments and her own internally designed tasks. The SBA tools included mental tests, homework, class work, assignments and projects, districts tests and class tests. Mary did not design formal examinations as this was done by the district. Although homework was used as a SBA tool, it did not feature as a prescribed assessment on the district work schedule term plan.

The district assessment plan listed the intended content for class work that should be assessed but Mary's SBA tools did not clearly track these specific topics. Learners' books indicated that class work was assessed on a daily basis with a few selected classwork tasks counting towards the learners' SBA marks. There were far fewer assessment tasks in the third and fourth term than in term 1. Mary did not articulate exact reasons for this but at the same time she indicated an expectation that common tasks would be distributed to the school, and suggested that this meant that she did not need to design additional tools individually.

d) Rules

Mary consulted the district work schedule to plan her assessments but did not always follow the rules and norms set out by the district. The district office had a stronger interventionist role than was visible in the S-school and the IC-school, determining the expected work coverage and pace as a pre-condition for district common examinations to be conducted at the end of each term. The district office constructed the common examinations in consultation with schools, but the actual topic areas were determined by external designers employed by the district. This placed pressure on the T-school and Mary to complete the required work as listed in the work schedule.

The district allowed the schools to construct their own individual class assessments but did not monitor Mary's compliance to their coverage rules. Mary's SBA tasks were mainly extracted from textbooks but some were constructed with reference to her background experience.

e) Community

As a result of high level district involvement in Mary's assessments, there was low school engagement among teachers on assessment tasks and moderation. The key tasks for the school were the common examinations. The moderation and feedback structures at the school level were mainly informal with a growing dependence on the district facilitators to mediate the assessment requirements and check for compliance. While the total number of tasks Mary designed was in line with the district expectations, the number of individual assessment tasks was substantially lower than in Kalay and Fiona's schools.

During the interviews, Mary indicated that she was comfortable with both the role the district was playing in facilitating common assessments and the information they received from the district facilitator responsible for Mathematics. She considered the guidance supplied by the district teams as necessary and valuable in upgrading her assessment skills. Mary indicated that the district held regular term meetings to bring schools on board with the externally set district common papers, and commented that she was comfortable with their style and level of language, and would not have made any changes to these:

I: Would your learners now be ok with that kind of structuring of questions?
M: Ja, it's ok
I: Or would you be ok with your learners answering?
M: It's ok for the learners … But would you make any changes?
I: No changes.
M: You wouldn't make any changes? You would be happy the way…
I: I'm happy

Mary said that she valued the coming together of teachers at cluster meetings to discuss the expected requirements for planning and assessing but seemed unclear about whether the information gained assisted her in designing assessments.

f) Division of labour

As with the S-school and the IC-school, the internal policies of the T-school determined the workloads and job allocations of individual teachers. Further, it was compulsory for teachers to plan and assess according to the district supplied schedules. While internal class based assessments were checked, no formal moderation structure within the school was evident. The management of the school used the district mediated documentation to monitor progress and compliance against district-level norms.

g) Summary activity system table for Mary

Table 6.7 below summarises the EMT analysis for Mary in the T school.

Table 6.7: Mary's SBA activity system

EMT Model	Assessment practice in the T-school

Subject	• Relatively inexperienced in teaching Mathematics at the primary school level. • Values district intervention at the school.
Object	• Mary was subject to a common district work schedule with district mediated goals • Her SBA tasks were meant to be standardised according to the common district work schedule but documented evidence showed that the district mediation did not always lead to Mary following the district schedule.
Tool	• In Mary's assessment plan, there are 7 different types of tools utilized. • Tools in terms 2, 3 and 4 are far fewer than those used in term 1. From term 2 onwards there seems to be a decline in the number of assessment tools and the only SBA tool used is the district common papers. • The district assessments have the highest overall weighting and are written by learners at the end of the term.
Rules	• It is compulsory for teachers at the school to cover the required work listed in the district schedules. • Mary follows the district plan up to term 1 and then follows her own plan without providing a justification for diverting. • Mary's assessments did not always align to the rules set out in the district work schedule.
Community	• The work of the district is prominent and overshadows the internal assessment structures of the school. • The district documents overshadow any direct dealing by Mary with national policy documents. • The school community is less involved in moderating and providing feedback to Mary on her SBA tools.
Division of Labor	• Mary is comfortable with the district arrangement and with her making little input into the process. • She is responsible for her own class based assessments. • The assessment work between the district and Mary is disjointed due to a lack of monitoring and engagement between the two parties.

h) Contradictions for the T-school

Primary contradiction: Division of labour

The provision of district assessment tools was associated in the T-school with limited contribution and ownership at school level to assessment activity. The high district mediation was prominent and overshadowed the internal assessment structures of the school. This contradicted the expected norms of teachers within the OBE approach (described in Chapter 2) for them to be effective assessors of learner competencies and "have the overall responsibility to assess the progress of learners

in achieving the expected outcomes" (DoE, 2005b, p.7). Lesh and Lamon (1992) have suggested that when teachers have more responsibility for assessment, assessment can truly become almost seamless with instruction; without this, they argue that there is a real risk of the two becoming separate entities with different objectives.

Although Mary valued the support and interventions facilitated by the local district office it had reduced her involvement in designing formal assessments and increased her dependency on receiving assessment tasks that are externally mediated and prescribed. Evidence in this study indicated that Mary designed and conducted fewer assessment tasks than Kalay and Fiona. This suggests that Mary became more dependent on an external supply of such tasks.

b) Secondary contradictions: Object $\longleftrightarrow$ Rules $\longleftrightarrow$Division of Labour

The rules of the engagement required Mary to complete the work as listed in the common district work schedule and thereby prepare learners for the district based common examinations. However, the link in Mary conducting her own SBA tasks and the preparation for the external district based common examination was not explicit in the district documentation. Teachers like Mary are left to their own understanding of what the purpose of the district interventions are.

6.5 Considering the mathematics seen in assessment artefacts

In this section, a detailed analysis of mathematics test questions (items) within the SBA activity system of each teacher and their respective schools is presented. Focus is placed specifically on the nature and selection of mathematics through an analysis of the SBA tools used by the subjects (Kalay, Fiona and Mary). The analysis unpacks the rules they applied to the selection of test items with comparisons made to the advocated design feature rules of the RNCS.

The discussion is based on the design features detailed in Chapter 4 and is illustrated with data and findings here. As in Chapter 5, the themes of mathematical coverage and the range and scope of questions again feature. Item maps were generated to analyse the mathematical coverage and the range and scope of questions. Bar graphs were used to represent the data and to show comparisons.

The assessment of common topics was added to compare the selections of mathematics in more detail. While the analytical commentary on mathematical coverage took into account findings from the overall SBA programme, the primary data source for the range and scope analysis was the final test/examination given to learners at the end of the academic year, since this assessment task was best suited for comparison with the ANA, allowing for commentary relating to key SNA and SBA artefacts. The data source for the third 'common topics' theme consisted of assessment tasks designed by teachers on selected common topics. Summary points are presented at the end of this section.

6.5.1 Mathematical coverage

The discussion on mathematical coverage was divided into two areas: 1) summaries of coverage and 2) the analysis of coverage based on categories relating to the overall SBA programme and final examinations.

Summaries of coverage

Since the overall SBA of teachers comprised a full academic year, it was relevant to establish the extent to which the different mathematical domains (LOs) and ASs listed in the RNCS were covered in each teacher's SBA. Analysis of coverage of mathematical domains established the nature and spread of LOs and ASs that featured in the teachers' assessment tasks, allowing for comparison with the weightings of these elements stipulated in the design features of SNA. This also allowed for coverage comparisons of domains between teachers' SBA and the expected requirements of the RNCS.

In the first part of this chapter I noted that all 3 schools met the policy requirement of at least two assessment tasks per term and a total of eight tasks for the year (DoE, 2003). In this part of the chapter analysis focuses on the specificity of curriculum coverage that was included in these assessment tasks. For each teacher, the information is summarised in two sets of tables (see **Table 6.8** through to **Table 6.13**) that provide a breakdown on the nature and spread of mathematics covered, by term, task, LO, AS, mathematics topic and mark allocation.

The first set of tables (**Tables 6.8, 6.10 and 6.12**) lists the LOs and ASs in focus within each assessment task (T) and the specific mathematics topics assessed within them. The numbering of the AS is linked to the LO as in the structure of content listed in the RNCS (DoE, 2002). The aim was to provide an overview of the LOs, ASs and topics selected within SBA by each teacher. Later in the analysis, I focus on the nature of test items that the teachers included within each LO in their design of formal examinations. For example Kalay had designed 10 assessment tasks and in the first task (T1) she selected only LO 1 and assessed work on 5 ASs (AS 1.2, AS 1.3, AS 1.4, AS 1.5 and AS 1.8) and 6 topics (counted by sub-bullets). Each teacher's mark allocations are included for completeness, rather than for analysis, and this matches to the RNCS advice that coverage weightings are linked to broad time proportions for each LO and the LO/AS/item level analysis of ANA in the SNA system. District assessments only featured in the SBA tasks of the township school.

In the second set of tables (**Tables 6.9, 6.11 and 6.13**) a summary of the number of ASs teachers selected in each task is listed per LO per term. In these tables distribution counts of individual ASs for each of the five LOs are indicated and the total number of individual ASs picked over the course of the academic year is indicated in the last row. This was done by looking at each question in an assessment task and linking it to a LO, AS and mathematics topic. Repeated counts of LOs were considered to show the frequency and spread of coverage across the academic year. For example, Kalay had selected LO 1 a total of 24 times across her SBA tasks. The combined total of her LO 1 to LO 5 selections was 54. In the analysis her frequency of selecting LO 1 was converted to a percentage (44%) and compared to the expected weightings of LOs documented in the RNCS (DoE, 2003a). A first and second set of tables is presented sequentially for Kalay, followed by Fiona and then Mary.

a) Kalay

Table 6.8: The mathematics selected in Kalay's overall SBA

Term	Task No.	LO	AS	Mathematics topic	Max. Mark

1	1	1	1.2 1.3 1.4 1.5 1.8	• Mental test: Number history - Properties of numbers • Mental Test: Natural and Whole Numbers - rounding off - factors and multiples - prime numbers • Mental Test: Fractions - Equivalent fractions - Addition of fractions	40
	2	1 and 2	1.4 2.3	• Place value • Flow diagrams - Input and output values	70
	3	1, 2 and 3	1.2, 1.3, 1.4, 1.8, 2.1, 2.3, 2.5, 3.2	• Control test 1 o Mixed operations o Rounding off o Place value o Flow diagrams o Fractions ▪ Brainteasers ▪ Word problems o Polygons	80
2	4	1 and 4	1.6 4.1 1.8	• Financial mathematics - Shop keeping problems - Calculating discount • Time - Time zones - Word problems	65
	5	4, 3, and 1	4.4 3.8 1.8	• Measurement - Working with Mass and Capacity - Length - Temperature • Grid work - Enlargements and reductions - Builds and investigates 2D and 3D shapes • Multiplication and Division - Short multiplication and division	60
			1.3 4.5	• Control test 2 - Multiplication and Division - Length - Capacity - Problem solving involving measurement	
3	6	3, 2 and 1	3.4, 2.1 2.4, 2.5 1.7, 1.8	• Rotational symmetry and transformations • Number patterns • Rate and Ratio • Multiple operations	60
	7	1	1.8	• Decimal fractions - Value - Conversions - Addition and subtraction	30

	8	1	1.5 1.8	• Control test 3 - Working out fractions of a quantity - Improper fractions and mixed numbers - Adding and subtracting of fractions	30
4	9	5 and 1	5.4 5.6 5.8 1.8	• Data handling and graphs - Bar graphs and tallies • Probability • Percentages	65
	10	3 and 4	3.1 3.2 4.8	• Angles • 3D shapes • Perimeter and area	50
	11	1, 3 and 5	1.3 1.8 3.8 5.8	• Summative assessment - Number operations - Fractions - Grid work and shapes - Probability - Percentage	45
		1, 3, 4 and 5	1.3, 1.8 3.4, 3.6, 3.8, 4.7, 4.8, 4.10, 5.5, 5.7, 5.8	• Control test – Term 4	55

Table 6.9: Kalay's AS selections within each Task

Term	Task	LO 1	LO 2	LO 3	LO 4	LO 5
1	1	5				
	2	1	1			
	3	4	3	1		
2	4	2			1	
	5	2		1	2	
3	6	2	3	1		
	7	1				
	8	2				
4	9	1				3
	10			2	1	
	11	2		1		1
End of year	CT	2		3	3	3
Total	12	24	7	9	7	7

Table 6.8 shows that all tasks except task 1 and task 7 tested more than one LO. Further, while the focus in task 7 was on only one AS (AS 1.8) the task still dealt with multiple skills within this AS. Also there was extensive revisiting of ASs in assessments across the year with the exception of ASs in LO5 which were met only once, in term 4. Each task comprised of sub assessment activities that contributed towards the total mark for an assessment. A total of 34 individual topics from 12 tasks were noted. **Table 6.9** shows a much higher prevalence of assessments

related to LO 1 ASs (24) than any of the other LOs.

b) Fiona

Table 6.10: The mathematics selected in Fiona's overall SBA

Term	Task No.	LO	AS	Mathematics topic	Max. Mark
1	1	1 and 2	1.2 1.3 1.4 1.8 2.5	• History of numbers • Number words and expansion • Place value • Rounding off • Addition and subtraction of whole numbers • Number sentences (speed and accuracy)	60
	2	3	3.1 3.2 3.3	• Identification of angles - Naming and measuring of angles - Identifying and measuring angles • Data handling - Pictographs and bar graphs	40
		5	5.1, 5.2		
		1 and 3	1.3, 1.4 1.8, 3.1, 3.2	• Cycle assessment on term 1 work	25
2	3	1	1.8 1.9 1.10	• Multiplication - Terminology • Multiplication and Division - Long multiplication and division - Word problems - using a calculator	100
	4	4 and 3	4.1 4.2 4.4 4.5 3.4 3.7 3.8	• Time – conversions • Time – word problems • Measurement – length, mass and capacity • Transformations • Lines of symmetry • Finding coordinates on a grid • Plotting coordinates on a grid	160
				• Mid-year Exam (term 1 and 2)	25
3	5	1 and 5	1.6 5.2 5.4 5.7	• Money – contextual problems • Data handling - Bar graphs • Reading and Interpreting graphs	50
	6	1 and 2	1.3 1.5 1.8 2.5	• Common fractions - Recognizing fractions - Conversions and calculations - Word problems - Addition and Subtraction • Decimal fractions - introduction - conversions - adding and subtracting • Number sentences (speed and accuracy)	80
		1	1.3, 1.5	• Cycle test on term 3	20

		and 5	1.6, 1.8 5.2, 5.4, 5.7		
4	7	1	1.3 1.4 1.7	• Decimal fractions - Recognition - Conversions - Expanded notation - Place value - Addition and subtraction • Rate and ratio	85
	8	4	4.4, 4.8, 4.9	• Length, mass and capacity • Perimeter and Area	73
		1 and 4	1.3, 1.4, 4.8	• Revision test	
		1,2, 4 and 5	1.3, 1.4, 1.6, 1.8, 1.7, 1.10, 2.5, 4.5, 4.7, 4.8, 5.7	• End of year examination	40

Table 6.11: Fiona's AS selections within each Task

Term	Task	LO 1	LO 2	LO 3	LO 4	LO 5
1	1	4	1			
	2	3		3		2
2	3	3				
	4			3	4	
3	5	1				3
	6	4	1			3
4	7	3				
	8	2			3	
End of year	Exam	6	1		3	1
Total	9	26	3	6	10	9

Table 6.10 shows that there were 39 topics assessed, the most of the three teachers. Tasks 3 and 7 were the only tasks that focused on one LO, and both of these worked across multiple ASs and sub-skills. A selection of the ASs was also re-assessed in her mid-year and the end of year examinations. Overlapping with Kalay, **Table 6.11** shows a much higher prevalence of assessments related to LO 1 ASs (26) than any of the other LOs. Fiona's SBA comprised 9 tasks and 39 topics with a combined total of 54 ASs selected across the LOs.

c) Mary

Table 6.12: The mathematics selected in Mary's overall SBA

Term	Task No.	LO	AS	Mathematics topic	Max. Mark

1	1	1 and 3	1.2 1.3 1.4 1.8 3.2	• History of numbers • Writing numbers in words • Place value • Basic operations: add, subtract, multiply and divide • Properties of triangles	108
	2	1, 3 and 5	1.3 3.1 5.4	• Fractions - Shaded parts of a whole - Decimals and the number line • Identify and name 2D and 3D shapes - Faces, vertices and edges • Frequency tables and data	105
		1	1.9, 1.3, 1.4, 1.6, 1.8, 1.10,	District assessment (DA) 1: mental calculations, number order, number words, place value, rounding off, expanded notation, comparing numbers, factors and multiples, addition and subtraction, long division, word problems, and products.	**50**
2	3	1	1.4 1.6 1.8	• Rounding off • Solving word problems on money • Fractions - Simplification and mixed fractions - Converting to decimals	140
		1, 2 and 3,	1.2, 2.1 1.4, 1.8, 3.2	DA 2: history of numbers, number patterns, number line, place value, number words, rounding off, addition and subtraction of whole numbers, subtraction of fractions, equivalent fractions, 3D shapes.	**50**
3	4	1, 3 and 4	1.3 1.7 3.1, 4.10 3.4	• Number words • Ratio • Naming angles and 3D shapes • Lines of symmetry • Reflections and translations	55
		1 and 2	1.3, 1.4, 2.1, 2.5, 1.6, 1.8	DA 3: place value, number words, number sentences, number patterns, rounding off, decimal value, mixed and common fractions, addition and subtraction of fractions, percentages, problems with money.	**50**
4	5	1, 2, 3 and 4	1.3 1.7 1.8 2.5 3.2 3.4 4.10	• Expanded number notation • Ratio • Rounding off • Number sentences • Constructing 3D shapes • Reflections and translations • Lines of symmetry • Classify triangles	50
		1, 2, 3, 4 and 5	1.7, 1.9, 2.3, 2.5, 3.2, 3.8, 4.5, 4.6, 4.10,	**District assessment 4**	50

			5.8, 5.9		

Table 6.13: Mary's AS selections within each Task

Term	Task	LO 1	LO 2	LO 3	LO 4	LO 5
1	1	4		1		
	2	1		1		1
	DA 1	6				
2	3	3				
	DA 2	3	1	1		
3	4	2		2	1	
	DA 3	4	2			
4	5	3	1	2	1	
End of year	DA 4	2	2	2	3	2
Mary's total excluding district assessments	5	13	1	8	2	1
District assessment total	4	15	5	1	3	2
Total	9	28	6	9	5	3

In **Table 6.12,** there are 26 individual topics assessed from 5 tasks excluding the district assessments. Because of the external supply of assessments, it became useful to look at Mary's coverage of individual ASs separately from the coverage seen in the district assessments. In **Table 6.13** Mary's coverage of ASs excluding the district assessments (23) was seen as significantly less than with them included (51). Mary own tasks incorporated very little coverage in LO 2 (patterns) and LO 5 (data handling), with district assessments providing some balance with their LO2 coverage in particular.

Analysis of coverage

The summaries of coverage allowed for comparisons at the level of number of tasks, LOs, ASs and number of topics. It is worth noting here ahead of the summary though, the decreasing number of assessment tasks as one works across socio-economic profiles in order from highest to lowest (S-school: 12; IC- school: 9; and T-school: 9 (5 + 4 district assessments) Also, in the township school, the number of topics covered in SBA across the Grade 6 year is lower than in the other two schools (26 topics noted in township school, compared with 39 for the inner-city school and 34 for the suburban school). Working at the level of AS I was able to represent the information of individual teachers graphically into groups of data for LOs and ASs, firstly for the overall SBA and secondly, for the final examinations in comparison to

the SBA tasks. These analyses allowed for comparisons on curricular emphases between the school's SBA with what was specified in the RNCS

a) The overall SBA

The statistics represented in **Tables 6.9, 6.11 and 6.13** on LO counts coverage per task were converted to percentages by taking the totals of each LO and dividing it by the sum total of all five LOs. Earlier I indicated that Kalay had selected LO 1 a total of 24 times out of a sum total of 54. This works out to 44%. Similar calculations were done for Fiona and Mary. Comparisons could then be made to the expected requirements of the RNCS. The expected weightings of LO indicated in the RNCS (DoE, 2003a, p.21) were 40% for LO 1 and 15% for LO 2, LO 3, LO 4 and LO 5 respectively. The overall SBA programme included the fourth term final examination coverage.

a1) LO Coverage

Figure 6.10 below shows the percentage weighting of LOs (with LO counts indicated on the right hand side of the graph) that Kalay, Fiona and Mary had in their SBA tasks in relation to the expected weightings of the RNCS. A relatively basic calculation is to exploratory measure "distance" from recommended weightings of the RNCS. For example, in Kalay's case differences were 12 points (4+2+2+2+2) from the recommended weightings.

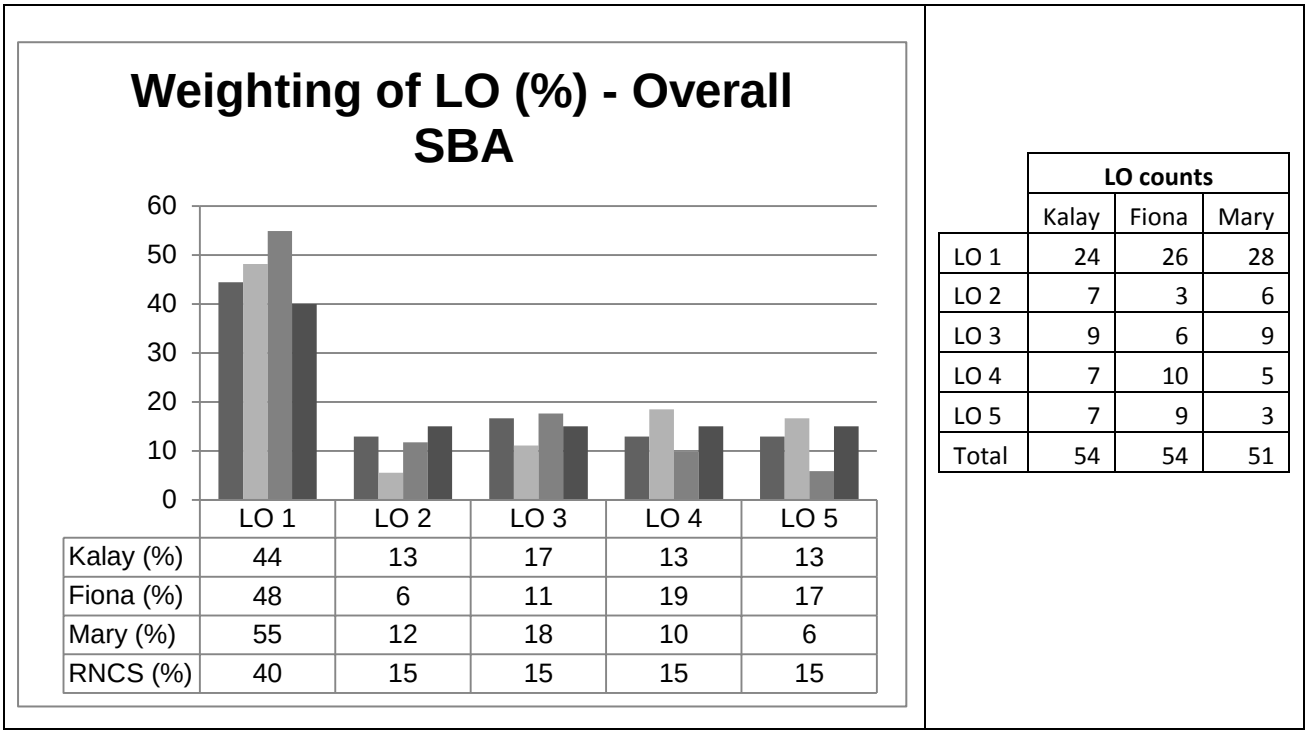
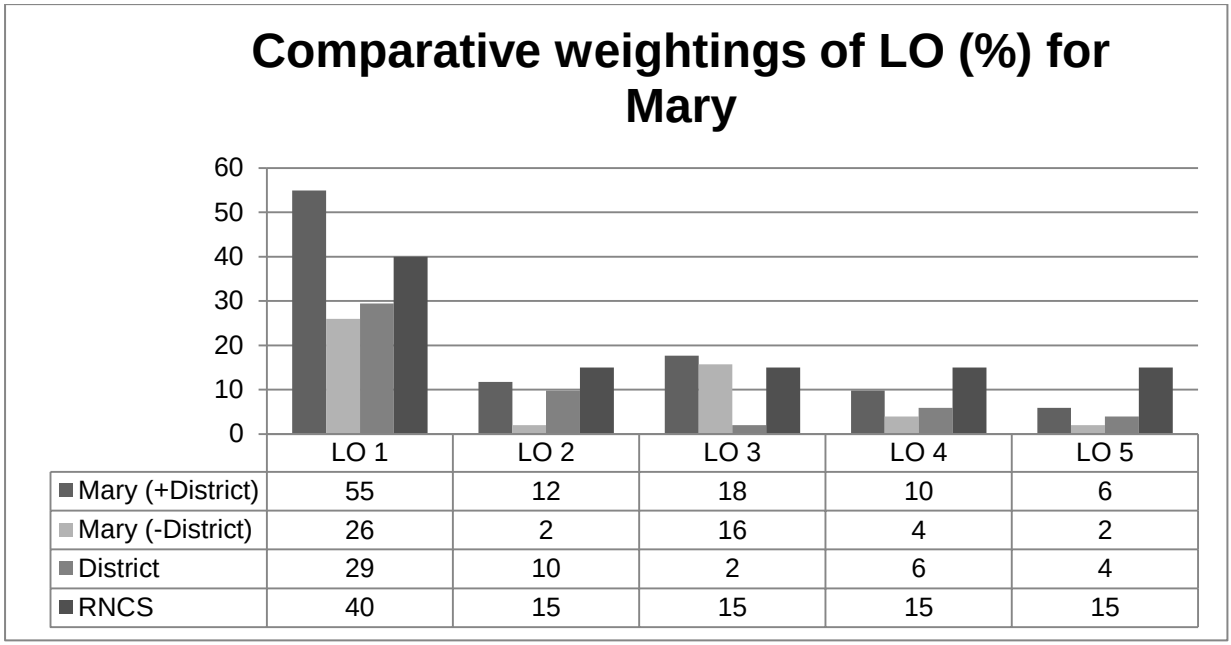


Figure 6.10: Percentage distribution of LO for the overall SBA

For Fiona, there was a difference of 25 (8+7+4+4+2) points from the recommended weightings, more than double that of Kalay (12=4+2+2+2+2). The statistics also show that Mary had the highest point's difference from the expected weightings (35=15+3+3+5+9) among the three teachers. This was mainly due to Mary's considerably higher than specified coverage of LO 1 compared to Kalay and Fiona. Across all three teachers there were misalignments with the expected curriculum requirements. For LO 1, Kalay, Fiona and Mary were above the expected 40% coverage. Kalay's weightings from the suburban school were the closest in alignment to the RNCS with marginal differences noted in LO 2 to LO 5. Fiona's and Mary's coverage across these LOs was erratic compared to the expected requirement of 15%.

The striking differences between Mary's coverage of LOs and the weightings of the RNCS prompted a further investigation into her self-designed assessment tasks and the district based assessments. The results showed that the coverage of work in the district assessments significantly affected the outlook of % weightings for Mary. In **Figure 6.11,** 4 distinct categories are listed for the T-school. Mary (+District) gives an account for the total LO spread in the township school inclusive of Mary's own

assessments plus the district assessments. This total LO spread is divided into subsets of Mary (-District) and District. Mary (-District) refers to the LO weighting seen in Mary's own assessments without taking into account the district assessments. District refers to the LO weighting observed when only district assessments are considered. The RNCS refers to the expected weightings from policy texts. The percentage weightings, except for LO 1, showed up significant differences between Mary's (-District) and the District weightings. Of interest here is that while the differences from the expected weightings are slightly reduced when the combined assessments across Mary and the district (+District) are considered, the balance of LOs across the district assessments considered separately (District) remain far removed from the expected weightings (difference of 49 percentage points = 11+5+13+9+11. This leads to the argument that district provision of assessments based on this dataset analysis is providing poor models of what expected coverage of LOs should look like.



	LO 1	LO 2	LO 3	LO 4	LO 5
Mary (+District)	55	12	18	10	6
Mary (-District)	26	2	16	4	2
District	29	10	2	6	4
RNCS	40	15	15	15	15

Figure 6.11: Comparative weightings for Mary

I cannot make claims about how Mary would have covered the LOs in the absence of the district assessments. It is clear that Mary's own designs (i.e. Mary (-District)) focused mainly on the coverage of 2 out of the 5 LOs (LO 1 and LO 3) with the other three largely ignored. It can be further noted that comparatively Mary's own designs has lower % coverage of LO 1, LO 2, LO 4 and LO 5 than Kalay and Fiona. The exception was LO 3, where she had a higher weighting than Fiona. The District also

emphasised LO 1 and LO 2 even more than Mary.

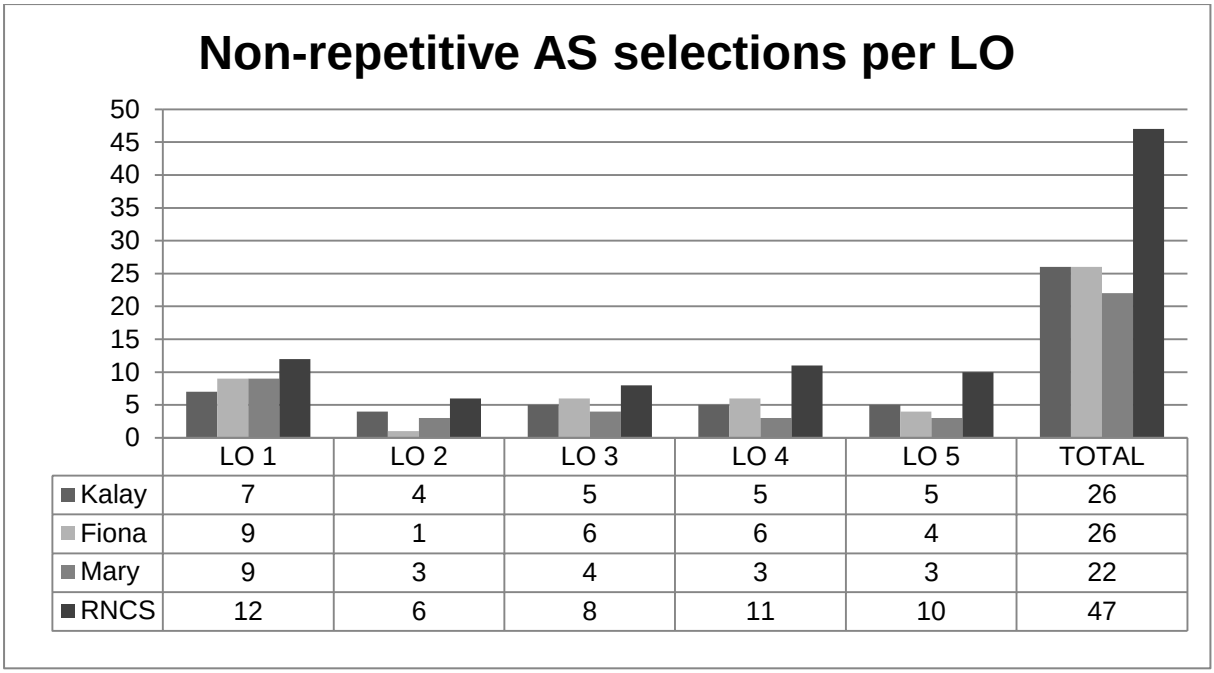
a2) AS coverage

Drilling further down than the LO level, I looked at the range of “checking in” at the level of ASs. Curriculum coverage as stated in the RNCS indicated that all ASs need to be covered in a year with emphasis on the fact that they represented minimal statements of attainment, and that it was important for learners to develop deep and meaningful understanding of them than to be rushed through them (DoE, 2003a. p.22).

> *These Assessment Standards are minimum Assessment Standards — that is, they show the minimum that a learner should be able to demonstrate at each grade level*.

However, the RNCS did not provide weightings for coverage of ASs so repeated counts of the ASs could not be compared to an expected percentage. Instead, from the structure of the ASs listed in the RNCS individual counts of ASs that were engaged with for each LO in SBA assessment artefacts could be determined. In Chapter 2 (**Table 2.8)** 47 ASs were listed in the mathematics curriculum across the five LOs in Grade 6 and a breakdown was provided on the number of ASs that featured in each LO.

At the AS level, the focus was on the range of individual ASs that were picked by the teachers in each LO rather than on the number of items a teacher selected on a single AS. Selections at the AS level were extracted from the data indicated in the AS column in **Tables 6.8, 6.10 and 6.12** listed earlier. Although in some tasks there was more than one item on a specific topic and AS, these were counted as ‘1’ instance of that AS and not taken as repeated counts. This mirrored the way I looked at breadth of coverage in AS terms in SNA in the previous chapter. For example in **Table 6.11**, Fiona had 26 single counts of ASs in her SBA programme. Fiona’s total number ASs selected was then compared to the total number of expected ASs (47) in the RNCS. Similar calculations were done for Kalay and Mary and these are shown in **Figure 6.12**, broken down by LO**.**



	LO 1	LO 2	LO 3	LO 4	LO 5	TOTAL
Kalay	7	4	5	5	5	26
Fiona	9	1	6	6	4	26
Mary	9	3	4	3	3	22
RNCS	12	6	8	11	10	47

Figure 6.12: Coverage of non-repetitive AS in SBA

I noted in the last chapter that it is not necessarily expected that all the content covered in the course of classroom teaching will necessarily be tested in SNA. The same can be said of SBA, even in the midst of policy documents calling for some assessment of all ASs. My focus was on a more exploratory and comparative investigation of the breadth of coverage of topics within SBA. **Figure 6.12** shows gaps for all 3 teachers in relation to AS level coverage with Fiona and Kalay covering less than 60% of ASs, and Mary covering less than 50%. The under-representation is spread across all LOs, but the gap differs across LOs and between teachers. For example, in Fiona's case, a total of 9 different ASs from LO 1 (75% coverage of this LO) featured in her SBA but there was an under-representation of ASs in LO 2 (patterns) (17% coverage of this LO) which only assessed learners' abilities to solve number sentences through speed and accuracy assessments. For Mary, there was an under-representation of ASs from LO 5 (data) which only assessed learners' abilities to solve frequency tables.

This implied that across all three teachers, there were a relatively high number of assessment standards that did not feature in their SBA programme. For Mary, the statistics indicated that more than half of ASs was not covered in her SBA programme, even with the support of district common papers. These statistics refer to assessment coverage and I do not infer from this, that these standards were not

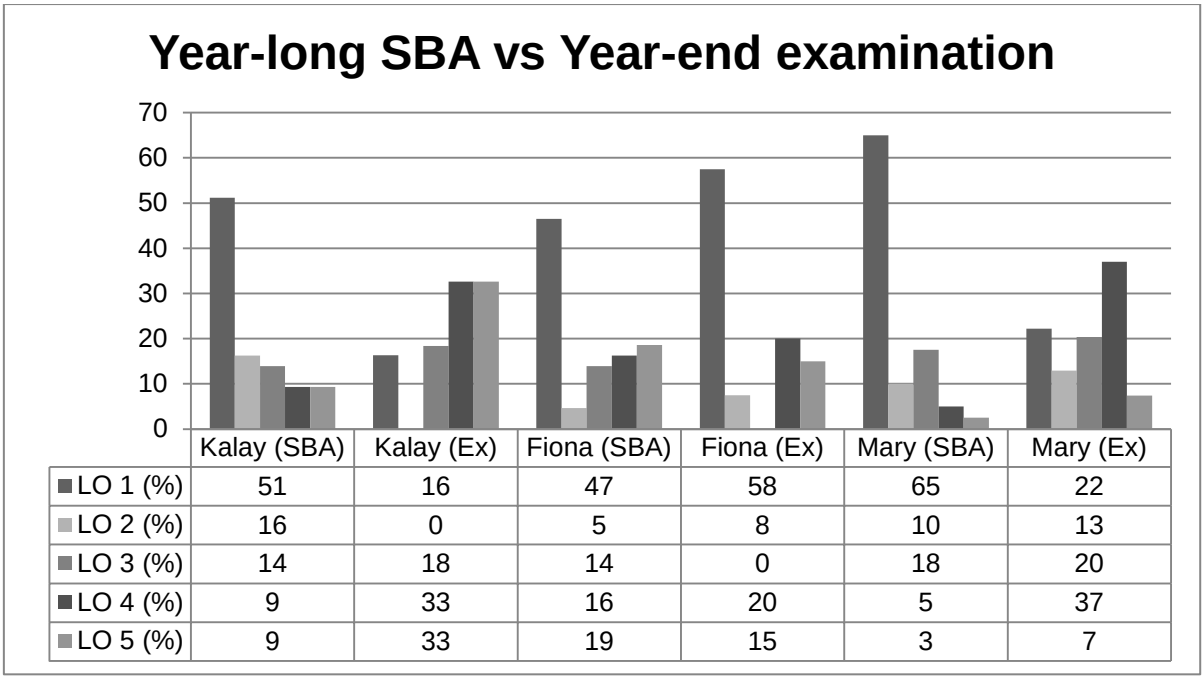
taught; rather that they were not assessed in the SBA programme.

b) Final examinations comparison with SBA

Although the RNCS did not compartmentalize content into termly schedules, the FFL assessment framework and milestones curriculum (DoE, 2008a) supplied by the DBE structured the coverage of all five LOs in each term according to the RNCS weightings indicated earlier. It was therefore a general expectation of curriculum monitors that teachers would similarly cover all five LOs in each term. The assessment plan supplied by the district for Mary had this structure as well, emphasising that the coverage of content in the year-long SBA tasks was to feed into the preparations for the range and scope of year-end examinations. It is was therefore useful to look at year-long SBA in relation to year-end SBA examination, on the premise that national policy promotes the idea that year-long SBA should feed into preparing learners for the coverage, range and scope seen in the year-end examinations, whether in the SBA or SNA system.

To get a more detailed account of mathematical content in year-end examinations, a similar approach to that used in Chapter 5 for analysing the ANA's using item maps was followed in order to establish mathematical coverage. Working at the level of individual items, an item map linking each item to an LO was created and then, using frequency counts, the data was grouped according to LOs. Percentages for individual LOs could be worked out from the total number of items appearing in the examination. These percentages were then also compared to the percentages of LO coverage in the year-long SBA tasks to explore whether the year-long SBA did indeed feed in and prepare learners comprehensively for the coverage, range and scope expected in the year-end SBA exam (and in later discussion, for the ANA papers in the SNA system).

Figure 6.14 below indicates comparative percentages on the coverage for each LO between the year-long SBA and the year-end examination (Ex) for each of the three teachers.



	Kalay (SBA)	Kalay (Ex)	Fiona (SBA)	Fiona (Ex)	Mary (SBA)	Mary (Ex)
LO 1 (%)	51	16	47	58	65	22
LO 2 (%)	16	0	5	8	10	13
LO 3 (%)	14	18	14	0	18	20
LO 4 (%)	9	33	16	20	5	37
LO 5 (%)	9	33	19	15	3	7

Figure 6.14: LO coverage – Year-Long SBA vs Year-end examination

In **Figure 6.14**, there are striking differences observed between SBA proportions through the year and proportions of the final examinations for all three teachers. For example, Mary's LO1 coverage in her year-end examination (which was the district assessment) is 22% compared to 65% in her year-long SBA tasks. There are also significant differences across all LOs. For Kalay and Fiona, there are significant differences in coverage of LOs between the SBA tasks and the examinations task. For Kalay, there is lower coverage in the examination of LO 1 and LO 2 but higher in LO 3, LO 4 and LO 5, inferring greater emphasis on content dealing with shape and space, measurement, and data handling in the final examination. Of the three teachers, Fiona shows closer alignment between the year-long SBA and the year-end examination on LO 2, LO 4 and LO 5.

The statistics in **Figure 6.14** also show striking differences in the coverage of LOs between the teachers' year-end examinations and the RNCS weightings shown in earlier graphs, indicating that teachers did not apply advocated weightings in their examination assessment tasks. Kalay gave more emphasis to shape and space (LO 3), measurement (LO 4) and data (LO 5). No questions on patterns (LO 2) were included. In contrast Fiona structured her examination to have 58% of the items on LO 1 and 20% of the paper on LO 4. With Mary, the district structured assessment showed deviations with the RNCS weightings on all five LOs. Overall then, two

results follow: firstly, there is limited match in all three school settings between the weighting across LOs seen in SBA tasks across the year and the year-end SBA examination; secondly, there is broad evidence that weightings on curriculum emphases are not being taken up in end of year SBA examination assessments, with the district assessment of Mary not able to close the gap. The coverage of mathematics by teachers in their SBA was followed by a more qualitative inquiry into the range and scope of questions the teachers utilised. This is discussed next.

6.5.2 Range and scope of questions

As explained in Chapter 4, the range and scope of questions selected by teachers were considered in relation to difficulty level (DL), cognitive demand (CD), language demand (LD), and item format (IF). The methodology for analysing the range and scope of questions using Item maps was explained in Chapter 4 and the list of item characteristics drawn from literature was presented in tabular form (see **Table 4.4,** p. 105) and illustrated through an ANA exemplar (see **Table 4.5,** p. 106).

In Chapter 5 these item characteristics were used to analyse three rounds of the ANA and data sets in the form of bar graphs were presented for each of DL, CD, LD and IF. In this section a similar coding approach is followed and item maps are used to generate these category data sets for Kalay, Fiona and Mary's SBA end year examination, enabling comparison between SNA and SBA artefacts that policy motives suggested ought to be structured by the same design features and weightings. From these item maps, item characteristics for each final examination item were generated. An exemplar data set for Fiona is indicated in **Table 6.14**.

Table 6.14: Exemplar item map for Fiona

Fiona		Question Number									
		1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10
	DL	E	E	E	E	E	E	E	M	D	D
	CD	K	K	R	R	K	R	K	R	C	R
	LD	LT	LT	LT	LT	LT	LT	LT	LT	LT	LT
	IF	CSA	CSA	CSA	CSA	CSA	CSA	CSA	CSA	CSA	CSA

From the item maps, data sets for each teacher in each category were represented alongside each other through bar graphs. The primary data source was again the final examinations given at the end of the fourth term. With this approach,

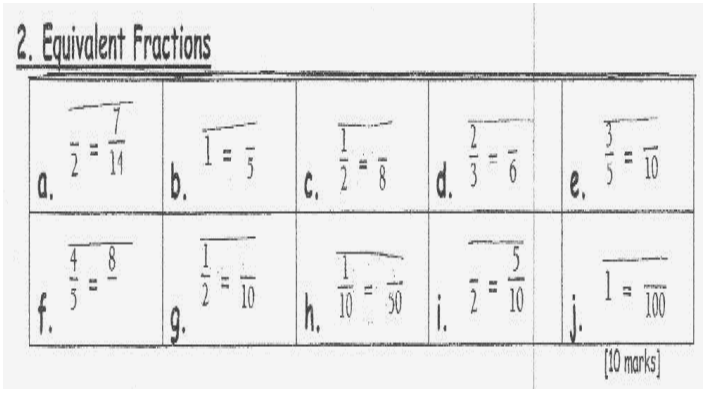
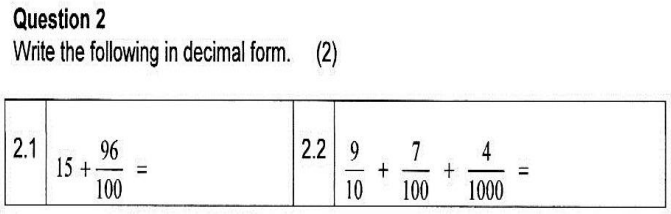
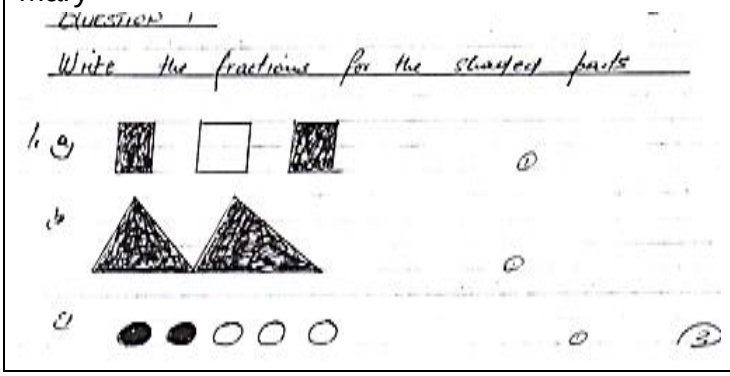
comparisons could be made with percentage breakdowns of categories indicated in the framework for the ANA as the SNA assessment artefact. Across each of the categories, the analysis was based on literature drawn descriptors used in the SNA analysis.

a) Difficulty level (DL)

Kalay stated intentions to design formal examinations with a balance of questions ranging from easy to difficult, and including familiar and non-routine questions. An analysis of the items in her end of year SBA examination showed that this balance did play through into her examination, where almost 60% of the questions were considered to be of moderate difficulty and familiar to learners (see **Figure 6.16)**. This balance was achieved even after her in-school community had cautioned her not to include too many high level questions because of the fear that the "results are going to be bad".

As in Kalay's case, the in-school community at Fiona's school wanted a fair balance of questions with the caution that if a test was considered too easy, it had to be re-done. Although Mary was comfortable with the standard and structure of the district paper, her learners still experienced difficulty in obtaining a pass mark. This suggested that Mary's learners were not meeting the higher grade-related concepts and skills in spite of the skewed presentation of items biased towards lower level items. Using the constructed categories for coding DL (Chapter 4): easy; moderate; and difficult that was linked to the curriculum content of a grade (Leong, 2006; de Lange, 1999), individual test items of teachers were matched and grouped. **Table 6.20** provides examples of my coding of difficulty levels of fractions-related questions drawn from all three teacher's tasks.

Table 6.15: Examples on difficulty level within LO 1 – Equivalent Fractions

Kalay  	Moderate difficulty: item linked to AS 6.1.8 in Grade 6 (calculates equivalent fractions).
Fiona  	Moderate difficulty: item linked to AS 6.1.8 in Grade 6 (calculates by selecting and using operations that involve common fractions).
Mary  	Easy: item linked to AS 4.1.3 in Grade 4 (common fractions in diagrammatic form). --------

Accordingly the results for the three teachers are indicated in **Figure 6.16**.

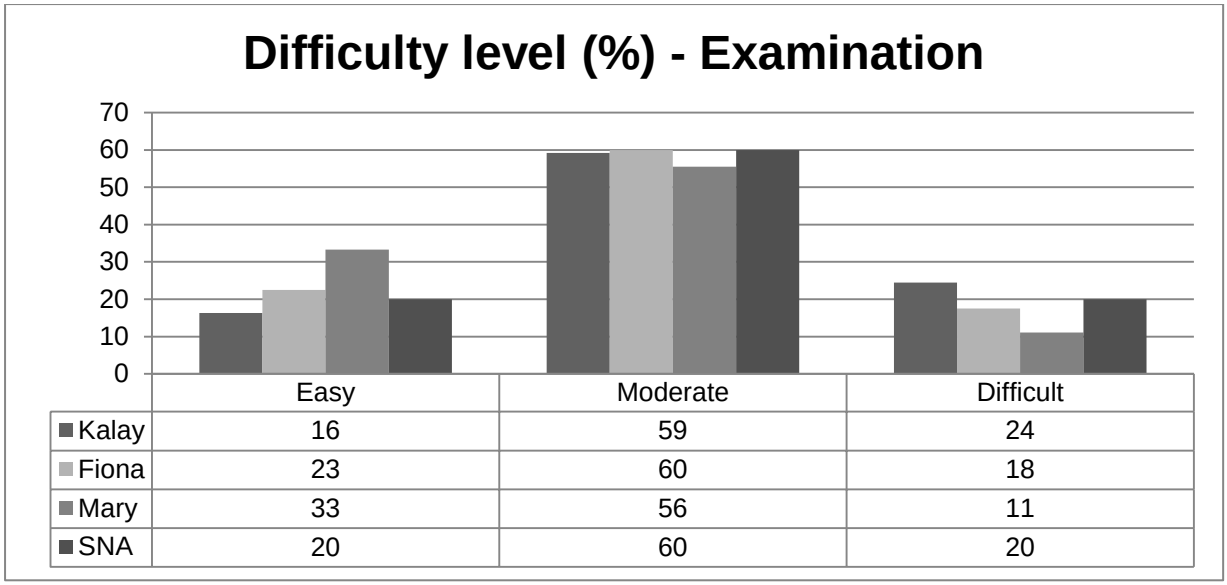


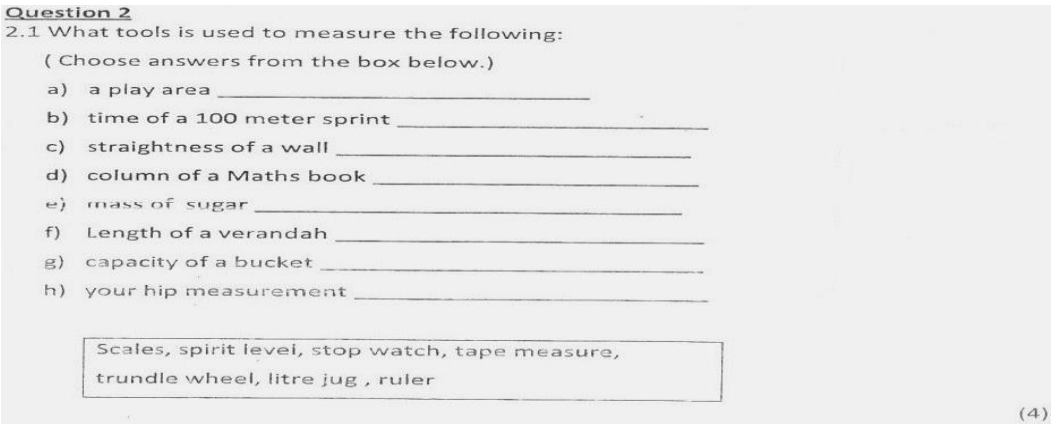
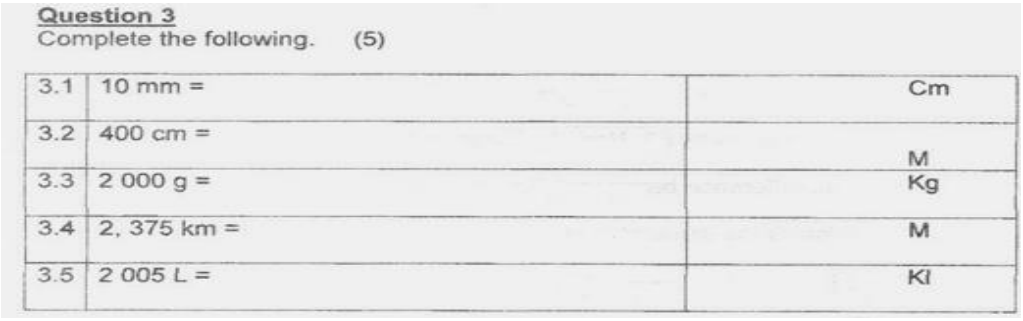
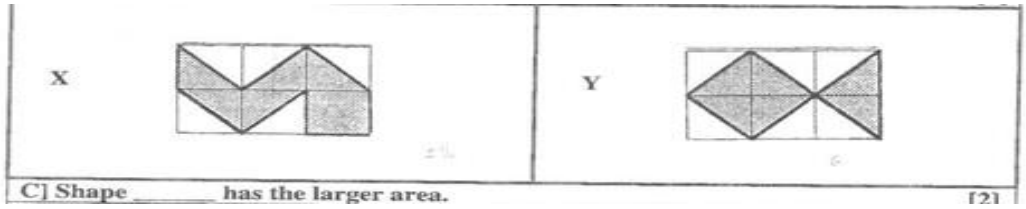
Figure 6.16: Difficulty level - Examination

The results in **Figure 6.16** suggest that the structure of the district assessment had a substantially higher percentage of easy questions (33%) and a lower percentage of difficult questions (11%) than Kalay and Fiona's assessments. The above analysis indicates that almost a quarter of Kalay's questions were interpreted as difficult and 16.3% were considered to be easy. Fiona's concern with the need for a mixture of easy, moderate and difficult question also played through in the above statistics with a DL breakdown close to the 20-60-20 split guidelines given to the test designers of ANA. In the T-school, the district-level offering of a common assessment had higher proportions of "easy" questions and the lowest proportion of "difficult" questions. While the two other schools largely reflected the advocated SNA DL emphases, the S-school showed the highest proportion of high difficulty items, and the T-school had a lower proportion than advocated of high difficulty items. Of interest though, in relation to the findings seen in Chapter 5 is that the the 11% proportion of high difficulty items seen in the T-school SBA examination reflects most closely the figure seen in the analysis of the 2008-10 ANA papers.

b) Cognitive demand (CD)

The CD descriptors used in the analysis were: knowing basic facts (K); applying routine procedures (R), using complex procedures (C), and solving non-routine problems. **Table 6.21** below illustrates typical examples extracted from the SBA tasks of teachers to exemplify the coding.

Table 6.16: Examples of CD coding

CD descriptor	Test item
Knowing basic facts (K)	**Item 1**  Question 2 2.1 What tools is used to measure the following: (Choose answers from the box below.) a) a play area _____ b) time of a 100 meter sprint _____ c) straightness of a wall _____ d) column of a Maths book _____ e) mass of sugar _____ f) Length of a verandah _____ g) capacity of a bucket _____ h) your hip measurement _____ Scales, spirit level, stop watch, tape measure, trundle wheel, litre jug , ruler (4) *Test item extracted from Mary.*
Applying routine procedures (R)	**Item 2**  Question 3 Complete the following. (5) 3.1 10 mm = Cm 3.2 400 cm = M 3.3 2 000 g = Kg 3.4 2, 375 km = M 3.5 2 005 L = Kl *Test item extracted from Fiona.*
Using complex procedures (C)	**Item 3** 4.2 Anne has 4 different skirts and 12 different blouses. How many different outfits can she make up using the 2 items? (a) 16 (b.) 48 (c) 8 *Test item extracted from Fiona.*
Solving non-routine problems (N)	**Item 4**  X Y C] Shape _____ has the larger area. [2] *Test item extracted from Kalay.*

Kalay indicated that the S-school afforded her the 'freedom to apply her creativity' to the design of assessment tasks. In this analysis, it was found that Kalay used varying combinations of scaling up and scaling down the number quantity and range

resulting in an increase or decrease in the cognitive demand of questions. Of the three teachers only Kalay included non-routine questions that targeted the skills of the higher aptitude learners and required a higher cognitive demand. Item 4 illustrates an example of a typical non-routine question. Kalay's proportions of non-routine questions were the highest among the three teachers.

There were two critical standpoints for Fiona in her design of the formal assessments that related closely to cognitive demand. Firstly she believed that the test papers should cater for learners from different ability groups (high and low aptitudes). She stated that this meant including questions of varying cognitive demand on the same topic in the same test or setting different tests. The option of setting different tests was not favoured by managers in the IC school. Secondly Fiona argued that her tests were not structured on how many marks the learners could get to pass but rather on opportunities for the learner to show evidence of understanding the concept being assessed. Fiona affirmed this in the interview.

M: So in a cycle test for example, you would have different types of questions.

I: Yes. Um…

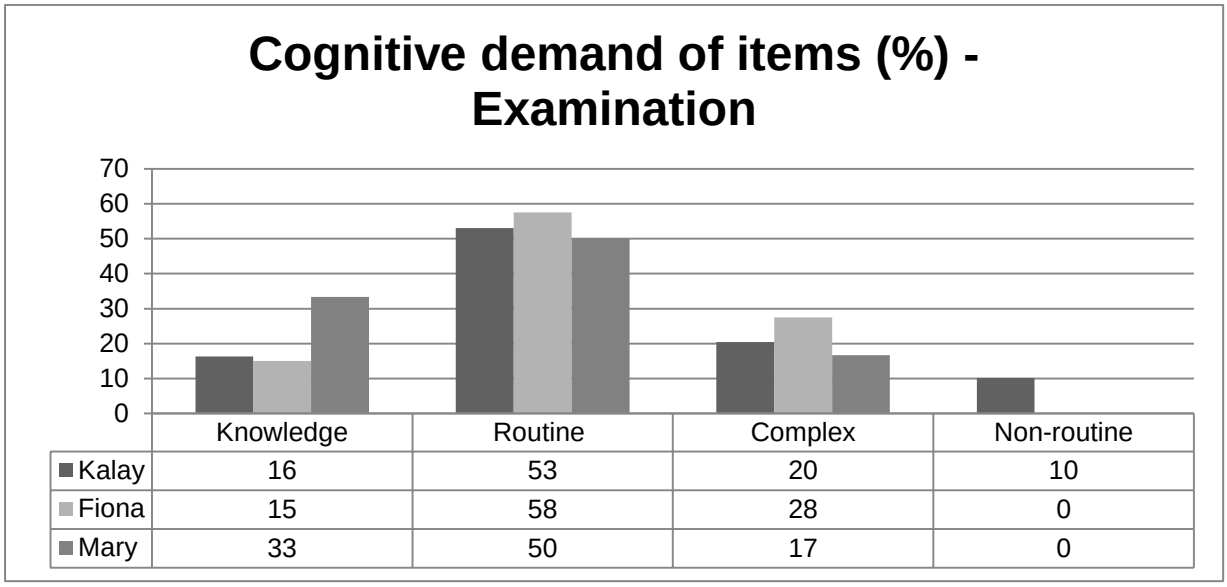
M: Like routine operations?

I: Yes, we've got…The way the paper is structured you'd find that it generally starts off with your basic knowledge questions – your routine questions rather, things that they should, you know, your bonds and tables are straight forward questions filling the missing answer. It develops to going on to where they have to now apply the knowledge that they have gained for those questions or whatever topics that was being discussed, going on to your problem solving skills.

Within this context, Fiona intended to design tests that were fair and balanced. This led to an inclusion of a high number of routine (item 2) and complex (item 3) questions.

The district common papers that Mary used were comprised of a variety of different types and styles of questions but none could be coded as non-routine. While the district did not supply a framework that the designers used for the district common papers, it was expected that they would be modelled in similar fashion to the ANA or

other SNA frames. The analysis though, indicated otherwise. As in the SNA analysis, the percentage spread of cognitive demand of questions closely resembled that of difficulty level. The results of the analyses for the three teachers are indicated in figure below.



	Knowledge	Routine	Complex	Non-routine
Kalay	16	53	20	10
Fiona	15	58	28	0
Mary	33	50	17	0

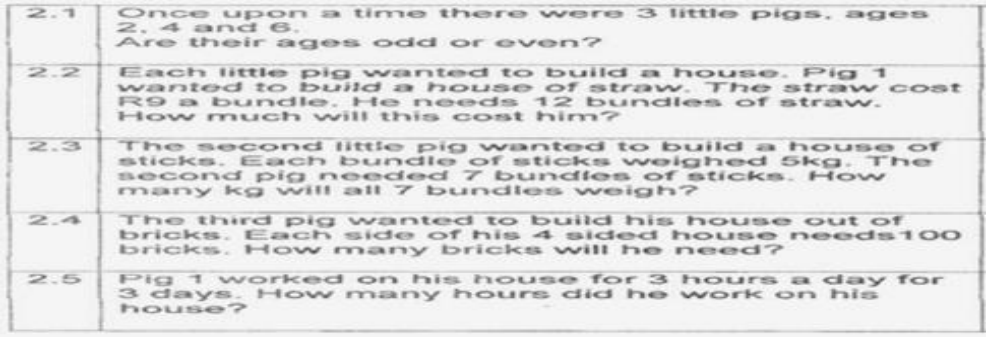
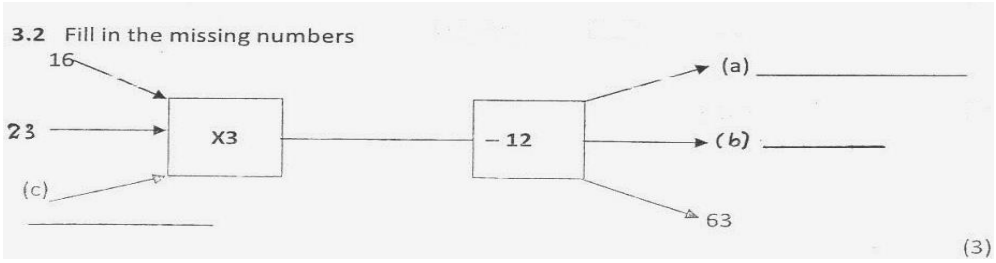
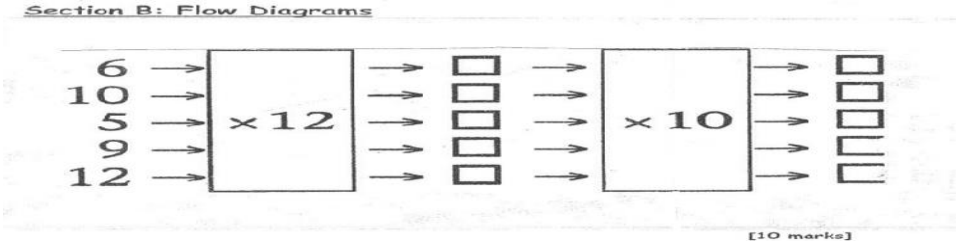
Figure 6.17: Cognitive demand (%) - Examination

Although the SNA policy prescripts did not stipulate a weighting for CD categories, the ANA (indicated in **Chapter 4**) targeted a 20% range for knowledge questions, 60% for application (combination of routine and complex) and 20% for non-routine problem solving. Of interest, **Figure 6.17** shows that all teachers were below the ANA stipulation for non-routine items and the T-school was above the stipulation at the lower cognitive demand end. The results showed a high percentage of knowledge questions (33%) for Mary in the T-school, and roughly double the proportion seen in the other two schools. Whether this was due to the district designing common assessments that would allow learners from underperforming schools to pass is unknown but as indicated earlier, school results suggest that this goal (psychological) was not being achieved. Fiona had higher proportions of routine and complex items than Kalay and Mary, but included no non-routine items. Overall on CD then, the patterns of spread across the levels followed the socio-economic profiles of the schools, with the most advantaged (S-) school's examination showing the highest push towards the higher levels.

c) Language demand (LD)

The LD descriptors used in the analysis were: high text (HT) or word problems, low text (LT), and no instructional text (NT). Items were coded as high text if they had more than one sentence in the stem and often included a context in the word problem. Low text (LT) items had instructions that were directed by short mathematical verbs (e.g. Add or Calculate). Items were coded no text (NT) when there was no instructions and the task was purely computational. **Table 6.17** below illustrates typical examples extracted from the SBA tasks of teachers to exemplify the coding.

Table 6.17: Examples of LD coding

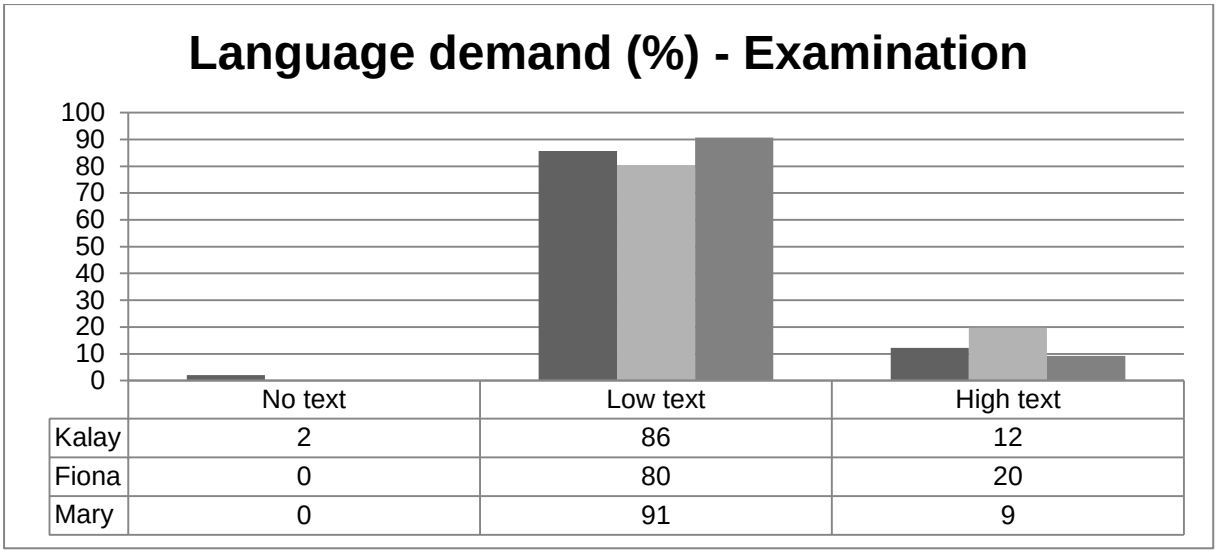
LD descriptor	Test item
High text (HT) (word problems)	**Item 1**  2.1 Once upon a time there were 3 little pigs, ages 2, 4 and 6. Are their ages odd or even? 2.2 Each little pig wanted to build a house. Pig 1 wanted to build a house of straw. The straw cost R9 a bundle. He needs 12 bundles of straw. How much will this cost him? 2.3 The second little pig wanted to build a house of sticks. Each bundle of sticks weighed 5kg. The second pig needed 7 bundles of sticks. How many kg will all 7 bundles weigh? 2.4 The third pig wanted to build his house out of bricks. Each side of his 4 sided house needs100 bricks. How many bricks will he need? 2.5 Pig 1 worked on his house for 3 hours a day for 3 days. How many hours did he work on his house? *Test item extracted from Fiona*
Low text (LT) (instruction/s is/are short mathematical verb/s)	**Item 2**  3.2 Fill in the missing numbers 16 23 (c) X3 – 12 (a) (b) 63 (3) *Test item extracted from Mary*
No instructional text (NT) (Computational)	**Item 3**  Section B: Flow Diagrams 6 10 5 9 12 × 12 × 10 [10 marks] *Test item extracted from Kalay*

A recurrent feature in Kalay's formal test designs was to exemplify the required skill and instruction with an introductory example. This device allowed her to clarify the

language instruction and provided learners with openings to make links to the given example and to recall similar work instructions done in class. Kalay regarded problems that had a context as ‘word problems’. In solving such problems, her expectation was for learners to handle a high amount of text by interpreting the meaning of the words, extracting relevant information, translating words into number sentences and symbols and finally applying a mathematical solution. Kalay indicated that she limited the weighting of high text problems in formal tests and examinations because learners did not fare well on them. This was confirmed in the analysis (see **Figure 6.18**).

Fiona noted the need for clear instructions to the learners who wrote and invigilators who conducted the tests. Therefore a set of instructions would appear (e.g. No calculators allowed) that would direct how she wanted learners to interpret instructions and show mathematical working. She, like Kalay, preferred low text questions in order to assess more in depth the mathematical operations that learners were exposed to in class. But she did consider it a necessary skill for learners (especially ‘brighter ones’) to be able to solve word problems (such as item 1)and the analysis showed that her assessment had the highest percentage (20%) of HT questions from the 3 teachers.

Mary indicated that she was comfortable with the levels of language used in the district common papers, and even if given the opportunity, would not have made any changes to it. There were some instructional differences observed between Mary’s formal assessments and the district end year assessment. Firstly, in some of Mary’s formal assessments tasks, no written instructions were provided and instead, oral or written instructions were given on the chalkboard. Secondly, the context provided would be explicitly high text or there would be no context at all. These are explained further in the analysis of common topics later in the chapter. While the results show that no NT questions appeared in the district assessment, these did feature in Mary’s internal formal assessments. The results of the analysis for the three teachers are indicated in the figure below.



	No text	Low text	High text
Kalay	2	86	12
Fiona	0	80	20
Mary	0	91	9

Figure 6.18: Language demand (%) - Examination

Across all 3 teachers the preference was to include questions that had low text, where a mathematical verb (e.g. Calculate) gave a clear instruction on what was expected. The proportion of low text questions ranged from 80% for Fiona to 91% for Mary's district assessment. The amount of high text questions (or word problems) ranged from 9% (Mary) to 20% (Fiona). Although the teachers made use of computational problems with no instructional text (NT) in their activity based assessments during the year (in particular Mary), there was no significant inclusion of these types of questions in the formal assessments. It was only Kalay that included NT items such as item 3 in examination tasks.

The results showed that the district assessment LD breakdown was similar to Kalay and Fiona's examination with the vast majority of questions categorised as LT. The proportion of LT questions (91%) in the district assessment was the highest of the three. In summary, while all schools showed similar preferences for low text questions, the S-school did most to scaffold learners into dealing with high-text items within the language format during the SBA activities through the year.

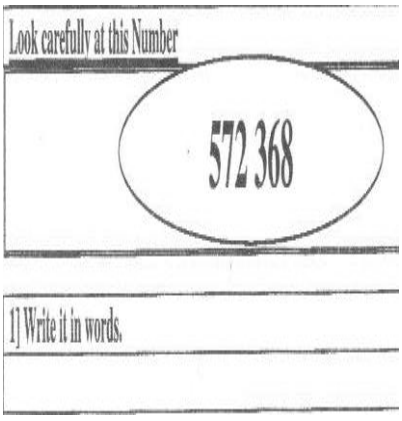
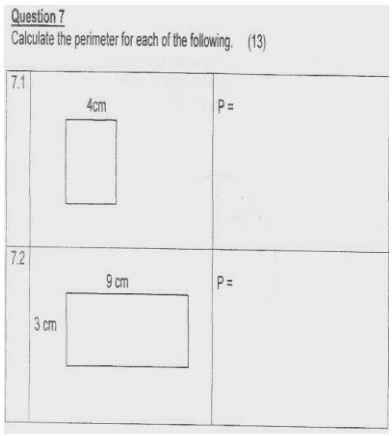
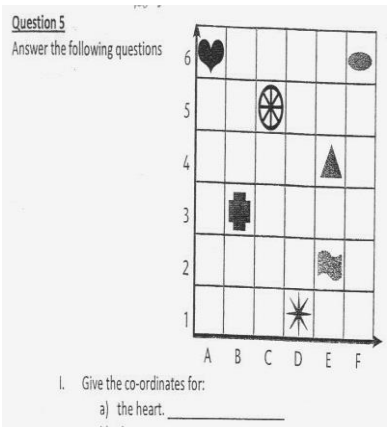
d) Item format (IF)

The item format (IF) descriptors used in the analysis were: multiple-choice questions (MCQ) or closed short answer response (CSA) questions. Items were coded MCQ when the learner was required to make a selection from a given number of distractors (see item 4). Under CSA items I incorporated a variety of short answer

types that included calculation response, writing down words, flow charts, tables, and grids. In the main, teachers preferred CSA type items to MCQ. Kalay's formal tests incorporated a range of representations across CSA questions that were drawn from various different textbooks and resources, which then, as a school rule had to be redesigned to be in suitable and accessible formats for learners and not just copied (see item 1).

Like Kalay, Fiona's CSA questions incorporated space for learners to complete short responses, calculate and show working, fill in tables and grids, work with geometric shapes, and construct and interpret graphs. Her test items were modelled on formats from a combination of textbooks and knowledge from past tests (see item 2). The district common papers (Mary) also contained a variety of different types and styles of questions (see item 3) but somewhat surprisingly, did not model the style of the ANA by including some MCQ items. **Table 6.18** shows the variety of items that were coded as CSA type.

Table 6.18: Examples of IF coding

CSA types		
Item 1 	Item 2 	Item 3 
Test item extracted from Kalay.	*Test item extracted from Fiona.*	*Test item extracted from Mary.*
MCQ types		

Item 4

4.3 Which number in place of # will make the number sentence true?
$14 \times \# = 280$

(a) 20 (b) 10 (c) 7

4.4 Which number in place of # will make the number sentence true?
$12 \div 3 = \# \div 12$

(a) 3 (b) 24 (c) 48

4.5 The fraction $5\frac{3}{4}$ is the same as :

(a) $5 \times \frac{3}{4}$ (b) $5 \div \frac{3}{4}$ (c) $5 + \frac{3}{4}$

Test item extracted from Fiona

Although the district common papers were used as reference for 'copying the style', Mary indicated that her principal source for designing her own formal tests were textbooks. There was a lack of alignment between the formatting of Mary's formal assessments and the formats used in the district assessment. In Mary's case, textbook questions were often directly photocopied onto test worksheets or re-written on the chalkboard for learners to copy and answer. The diagrams, though not drawn to scale, allowed learners to respond but there was often limited space for learners to show calculations or rough work if needed in Mary's own SBA tasks. The results of the analysis for the three teachers are indicated in the figure below.

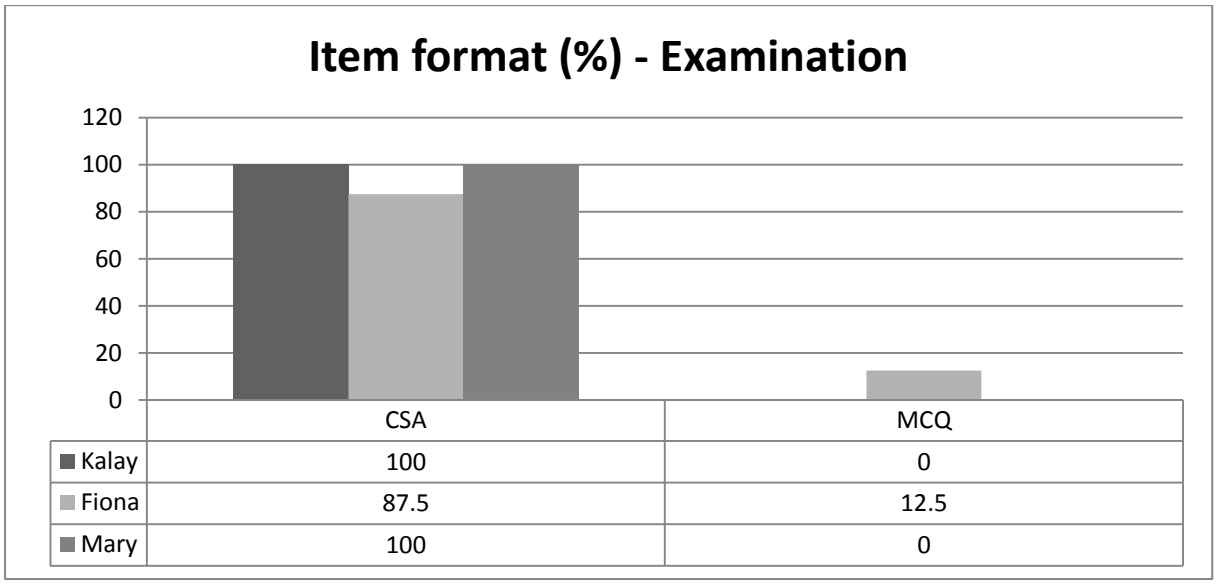


Figure 6.19: Item format (%) - Examination

Across the 3 teachers, the analysis showed an overwhelming preference by the

teachers for CSA type questions with Kalay and Mary's district assessment showing a 100% of this type. Only Fiona incorporated a spread of CSA (87.5%) and MCQ (12.5%) and could claim some resemblance to the IF formats used in the standardised ANA. There was however, variety in the way the teachers constructed their CSA type questions as a result of the different textbooks and resources used.

e) Summary

The DL, CD, LD and IF analyses showed teachers in different school settings applied different rules in designing mathematics items. Significant differences in percentages scores could be observed between Mary and the other two teachers on DL and CD. The coding on DL showed little alignment between the teachers and the ANA. Only Kalay had NT items and only Fiona made use of both MCQ and CSA item types. Comparative differences in the coding of DL, CD, LD and IF are summarised in the **Table 6.19**.

Table 6.19: Summary of the range and scope of questions

Category	**Descriptor**	**Teacher**		
		Kalay	**Fiona**	**Mary**
DL	Easy (E), Moderate (M) and Difficult (D)	Majority of items were moderate. She had the highest range of difficult items	Majority of items were moderate. Fiona's spread of E-M-D was close to that 20-60-20 spread recommended for the ANA.	Majority of questions were moderate. Mary had the highest % of easy items and lowest % of difficult items
CD	Knowledge (K), Routine (R), Complex (c), and Non-routine (N)	High amount of R and C type items with moderate amount K and N items. Only teacher to have non-routine problems classified.	High amount of R and C type items with moderate amounts of K type items. No N type items were classified.	The district assessments in Mary's case have a high amount of K and R type items. Moderate amount of C type items with no N type items classified.
LD	High text (HT), low text (LT) and no instructional text (NT)	Mainly LT – focus on abstract applications of laws. HT items used in complex problem solving	LT items – questions relatively free of contexts.	The amount of LT items (91%) in the district assessment was the highest of the three teachers.

		and non-routine. A small % of NT items was observed.		
IF	Multiple choice questions (MCQ) and closed short answer (CSA) questions	Only CSA type items classified. High number of questions extracted from textbooks and past year papers and redesigned for learners.	Combination of CSA and MCQ type items included but the majority were CSA type. High number of items sourced from textbook and re-printed.	Only CSA type items classified in district assessments. With internal assessments, learners mainly re-write questions from board. Almost all the questions extracted from textbook.

6.5.3 Assessment of common topics

The final theme of the SBA artefact mathematical analysis involved an assessment of common topics. Within this theme I considered each teacher's selection of mathematics in relation to common topics, and analysed these selections in terms of the categories used previously: mathematics coverage, difficulty level, cognitive demand, language demand, and item format. The data sources here were the SBA assessment tasks used to assess common topics and the second interview held with teachers where they were probed on their task selections. A critical purpose of this theme was to facilitate comparisons on design rules among the three schools when the topics were the same.

The following common topics were selected (see **Table 6.20**) as they satisfied key conditions: a) they appeared in the portfolio files of all three teachers, b) they covered three learning outcomes that dealt with different mathematical domains (e.g. number, measurement and data handling) and c) assessment activities on these topics were consistent with evidence in the learners' work books.

Table 6.20: Common assessment topics according to LO

Learning outcome	Topic
LO 1 - Number	Solving problems involving rate and ratio.
LO 4 - Measurement	Measuring and calculating the capacity of objects.

LO 5 - Data handling	Reading and interpreting graphs.

In the analysis, tasks were coded sequentially according to the teacher and the topic for ease of reference in the discussion. For example the first task chosen for analysis designed by Kalay on Rate and Ratio was coded as KR 1, the one designed by Fiona as FR 1 and the one by Mary as MR 1. A similar coding style was applied to the Capacity and Graphs tasks.

Rate and Ratio

a) Expected skills

The first common assessment topic selected for analysis involved rate and ratio (from LO 1). Within this topic, the curriculum specified that learners should recognise, describe and represent numbers and their relationships with competence and confidence in solving problems. The AS in the RNCS (DoE, 2002b) for Grade 6 was outlined as follows:

The learner will be able to solve problems that involve:

- *Comparing two or more quantities of the same kind (ratio);*
- *Comparing two or more quantities of different kinds (rate, e.g. wages/day)*

As discussed in Chapter 2, the AS specification in the RNCS does not pinpoint the exact skills to be assessed. Research studies on ratio and proportional understanding (Hart, 1981; Cramer, Post & Currier, 1993; Misailidou and Williams, 2003) have indicated that expected skills on this topic can include doubling and halving, multiplication by an integer, multiplication by fractions, application of rate, enlargement, percentage and patterns and relationships. Hart (1981) pointed out that the ratio test designed by the CSMS[8] team included items that required the child to draw, use numbers in relation to diagrams and presented in problems. Misailidou and Williams (2003) argued that questions on ratio can vary in representation across written algorithmic statements to models that involve pictures, tables or double lines. Their research revealed that learners did significantly better on ratio questions that had pictorial representation (55.2%) than when they were represented in other forms

[8] The CSMS team designed tests and conducted investigations on several mathematics topics among British school children aged 11-14.

(17.2%). Hart (1981) further pointed out that applications of ratio can involve knowing proportionality facts (knowledge), applying proportionality-related rules, and using the rules to solve problems.

Common strategies for solving rate and ratio problems include doubling and halving, finding the unit rate and using addition strategies (or enlargement). According to Hart (1981) the CSMS research revealed that ratio questions that could be solved by using doubling and halving were easiest when presented in either word problem or drawing form. Depending on specific number range, questions that involved finding the unit rate and understanding its meaning were more difficult to solve. However, Cramer, Post & Currier, (1993) in their research conducted in the Rational Number Project (RNP),[9] concluded that the unit rate was the strategy that generated the largest number of correct answers. The unit rate is found through division and represents the constant factor between the two quantities. The use of additive reasoning (scaling up by the addition of a small quantity), instead of multiplicative reasoning has been highlighted as a common source of error in ratio questions (ref).

Hart (1981) noted that the majority of children in their English study did not progress beyond doubling and halving in solving rate and ratio problems. This was often due to difficulties with multiplying by a fraction in solving more difficult ratio problems. Hart (1981) further mentioned that teachers should be aware in their design of assessment items that children avoided multiplication by a fraction and often used repeated addition.

> *The introduction of non-whole numbers into a problem does not make the question a little harder but a lot harder (Hart, 198, p. 101).*

Karplus referred to this phenomenon as the 'fraction avoidance syndrome (Cramer, Post & Currier, 1993). Misailidou and Williams (2003) in their research on ratio argued strongly that even experienced teachers were often not aware of the kinds of misconceptions that learners tend to exhibit.

Similarly, the RNP also revealed that numerical tasks with rates having non-integer relationships were more difficult for learners to solve than those with integers. This

[9] The RNP administered a survey of proportional reasoning tasks to over 900 middle grade students.

research also showed that comparing unequal rates was more difficult than comparing equal ones. Misailidou and Williams (2003) pointed out that this is not new phenomenon as far as ratio is concerned. Extended research from as early as 1966 (Lunzer & Pumfrey, 1966) has revealed that solving ratio and proportion problems is a very difficult task for most pupils in the middle school years throughout the world. Cramer, Post and Currier (1993) also suggested that a large segment of society never acquires well-founded understandings of rate and ratio at all.

This literature forms a brief background to the analysis of how the three teachers selected questions and tasks on the topic of ratio in their assessments.

b) Tasks

The tasks designed by the teachers exemplifying coverage on Ratio are indicated in **Table 6.21** below. KR 1 is two tasks selected from Kalay's assessment tasks on Rate and Ratio. FR 1 is an extract from Fiona's tasks and MR 1 was extracted from Mary's tasks. Mary's tasks, unlike those from Kalay and Fiona, were taken from a textbook that she used often in classroom activities that presented the entire task in the context of AIDS patients (see MR 1).

Table 6.21: Assessment tasks on Ratio

KR 1

TASK ONE [10 MARKS]

Rate

If 2 apples cost R5,00, what is the price of:
1 apple: (a) ..... 3 apples: (b) ..... 5 apples: (c) .....
500 g of rice costs R8,00. What is the price of:
1 kg rice: (a) ..... 3 kg rice: (b) ..... 2 ½ kg rice: (c) .....
400 mℓ of liquid soap costs R8,80. What is the cost of:
200 mℓ: (a) ..... 1 litre: (b) .....
2 1/10 litres: (c) ..... 4 ½ litres: (d) .....

TASK TWO [10 MARKS]

Ratio

The pocket money rule in the Sibiya family is that Mpho gets 3 times as much as Ntombi. The ratio is 3:1.
(a) If Mpho gets R3, Ntombi gets .....
(b) If Mpho gets R....., Ntombi gets R3,00.
(c) If Mpho gets R12, Ntombi gets .....
(d) If Ntombi gets R12, Mpho gets .....
(e) If Ntombi gets ..... Mpho gets R15,00.
(f) If Ntombi gets R4,50, Mpho gets .....
Mr Tema travels at a speed of 60 km/hour.
(a) How far would he travel in 3 hours? .....
(b) Mr Monisi travels twice as fast. What is his speed? .....
(c) Cyclist Andile travels half as fast as Mr Tema. What is cyclist Andile's speed? .....
(d) Mr Tema travels three times faster than cyclist Ben. What is the speed of cyclist Ben? .....

FR 1



Grade : 6____ Total : 10

LO: 1 AS : 6.1.7 Parents Signature :_____

Task 2 : Activity 1 : Rate and Ratio

1. There are 7 blue pens, 5 black pens, 3 green pens and 4 red pens.

Complete the ratio of : (5)

1.1	The number of green pens to the number of black pens	
1.2	The number of blue pens to the number of black pens	
1.3	The number of blue pens to the total number of pens	
1.4	The number of black pens to the number of blue, green and red pens altogether	
1.5	The number of green pens to the number of red pens	

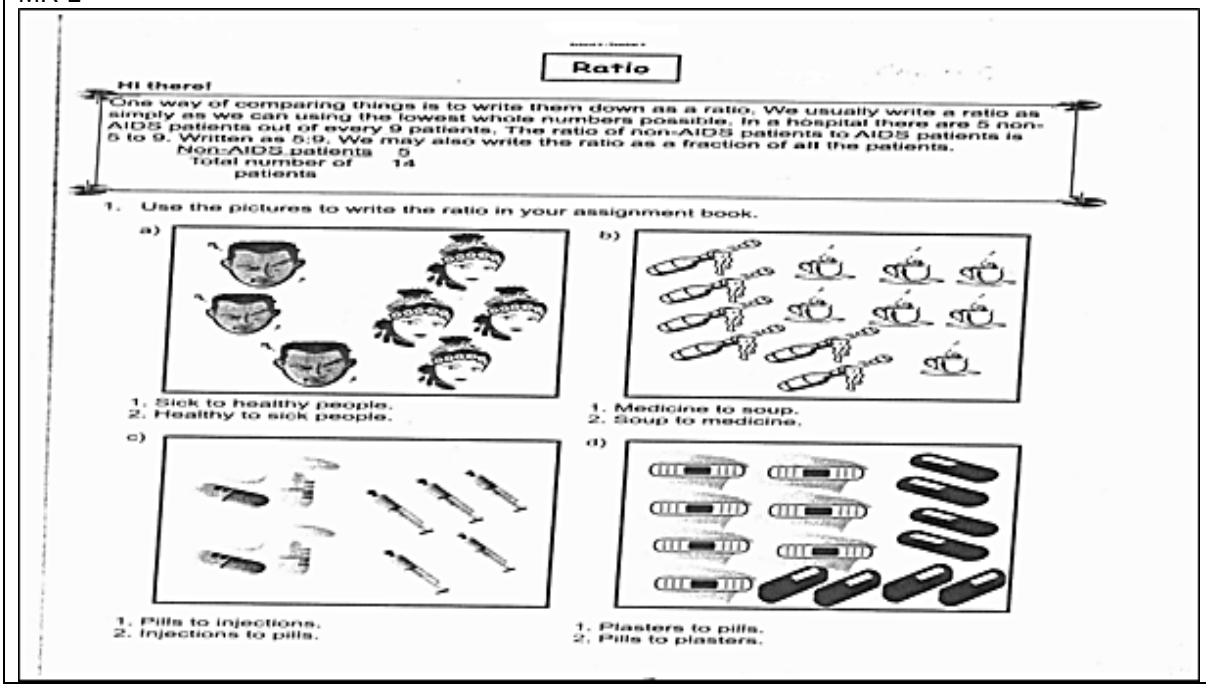
2. Recipes use a particular ratio of ingredients so that the amount of food or the number of cakes, tarts and biscuits can be varied without altering the taste or texture.

Ingredients for Ashleigh's crumpet recipe. (5)

200ml flour
100ml sugar
2 eggs
2 teaspoon baking powder
A little milk

2.1	Write down the first three ingredients if the recipe is doubled.	
2.2	How much sugar will be needed if 6 eggs are used?	
2.3	The ratio of the amount of flour to the amount of sugar is	

MR 1



Ratio

Hi there!

One way of comparing things is to write them down as a ratio. We usually write a ratio as simply as we can using the lowest whole numbers possible. In a hospital there are 5 non-AIDS patients out of every 9 patients. The ratio of non-AIDS patients to AIDS patients is 5 to 9. Written as 5:9. We may also write the ratio as a fraction of all the patients.

Non-AIDS patients 5
Total number of patients 14

1. Use the pictures to write the ratio in your assignment book.

a)

1. Sick to healthy people.
2. Healthy to sick people.

b)

1. Medicine to soup.
2. Soup to medicine.

c)

1. Pills to injections.
2. Injections to pills.

d)

1. Plasters to pills.
2. Pills to plasters.

c) Analysis

Coverage (MD)

Kalay covered both aspects (ratio and rate) of the AS and incorporated a range of calculation skills in both her tasks. KR 1 involved both scaling up and scaling down of quantities. In task 1, all questions involve 'graded' sub-parts with the first part always presenting a doubling or halving instance, described in the literature overview as easier for children (Hart, 1981). In FR 1, Fiona also designs the assessment to cover the two aspects of the AS from the Grade 6 curriculum. The assessed skill in the first question required identifying two parts of a ratio and parts in relation to the whole quantity. The second question involved the calculation of rate and included

comparing same and different quantities.

In MR 1, the order of the ratios was given prominence in all questions with the numbers already simplified. There was limited assessment of skills on the ratio AS and fewer in comparison to Kalay and Fiona. In Mary's case, the photocopied questions contained the use of relatively specialized terms in the text (e.g. injections, plasters) and some contextually unrealistic and potentially problematic associations between terminology and diagrams. There were both mathematical errors and ethical issues relating to this way of handling a sensitive topic such as AIDS in South Africa. There was also an error in the lead information since if the total number of patients was 9, then the number of non-AIDS patients to the number of AIDS patients should have been written as 5:4 and not 5 to 9.

Difficulty level (DL)

Using the defined DL indicators (easy, moderate and difficult) based on difficulty in relation to the curriculum grade requirements linked to knowledge dimensions (Leong, 2006), the difficulty level of questions for the three teachers was established. Kalay's and Fiona's tasks ranged from moderate to difficult since their tasks were in keeping with the ASs listed for Grade 6. It is worth noting that Kalay also indicated that calculating rate and ratio was not easy for learners.

K: It suits the… It caters for the different levels.
I: Like if you say 'If 2 apples cost R5 what is the…
I + K: … price of 1 apple'
K: So they must halve it. Here, ok I wouldn't say it's that easy.

KR 1 items that were coded difficult were those that involved a sequential increase in the calculations from working with whole number quantities to fractional quantities (2½ kg), and working across different measures (grams/kg) using ratio ideas. For example, question 6 in task 2 was coded as more difficult than the first five since it incorporated fractional values. In FR 1 the second question on recipes was coded as difficult (which Hart (1981) also referred to as generally more difficult). There is some interpretation of text ("If 6 eggs are used") and a second quantity ("the sugar") has to be worked out. For Mary, in MR 1, the questions were all regarded as easy as they were below the Grade 6 level AS. Simple skills were tested with little variation in the

examples. Only the (part: part) ratio is required and the prior knowledge required relates essentially to counting. Her DL levels are below the Grade 6 curricular requirement.

Cognitive demand (CD)

The coding applied here was based on the earlier defined CD indicators (knowledge, routine, non-routine and complex). In KR 1, the cognitive demand of questions overlapped between routine and complex procedures. In the two tasks, the application went beyond knowledge recall and a multiplicity of skills was assessed in each sub-question. In FR 1, question 1 is mainly centered on knowledge recall while question 2 involves routine problem solving. Question 2 involved a routine skill of solving for a single unknown in relation to known quantities. There is limited use of fractions and decimals in the questions to extend the cognitive demand of learners.

In MR 1, all the questions were at the knowledge level, and based on topics that learners would have been exposed to in previous grades i.e. referring to given information and recalling the symbolic representation of ratio and applying it to four diagrams. Misailidou and Williams (2003) research indicated that learners did significantly better on ratio questions that had pictorial representation than when they were represented in other forms as they were cognitively less demanding. In each of a, b, c and d, the ratio skills were restricted to two representations of ratio – the second representation being the reverse of the first.

Language demand (LD)

The defined LD indicators applied to the tasks were (No text, Low text and High text). In KR 1, there is inclusion of high text decimals and money applications to extrapolate the mathematical concepts tested. For example in KR 1-task 2, in questions 7 to 10, there is a range of mathematical texts that include questions where mathematical concepts of speed, distance and time are tested. With the inclusion of speed, Kalay shows the implicit presence of two variables and the learner is required to interpret the word problems through a critical reading of the text.

In FR 1 an everyday context (e.g. recipes) was used. According to the LD indicators,

both questions could be considered as high text problems. In the second question the lead information drew attention to the meaning of a recipe, in addition to the mathematical text of the problem, namely, “recipes use a particular ratio of ingredients so the amount of food or the number of cakes, tarts and biscuits can be varied without altering the taste or texture.”

In MR 1, The LD of the AIDS context was considered as a high text application. Firstly, the concept of ratio is introduced within the context of HIV and AIDS and learners have to link the given texts to numerical ratios. The text is misleading because of the error in the explanation of the part-whole concept although there is an example given. Secondly, in some of the questions, the terminology used in the questions was not consistent with that used in the contextual information. So learners had to make certain linguistic connections between ‘sick’ and ‘AIDS’ patients. Mary indicated no reluctance in including a sensitive context such as AIDS to assess learners on ratio.

M: They liked the ratio…They find the context, sometimes it stimulates them?

I: And is HIV and Aids not a sensitive area for them or are they…

M: They don’t mind? They know about it. These days everybody knows about…

The LD of MR 1 is high text but with low DL and CD levels

Item format (IF)

The IF indicators applied were closed short answer (CSA) and multiple choice questions (MCQ). In KR 1, the IF was CSA type showing calculations. There was indication of the marks and each task had an overall instruction. The format of the question was extracted from a textbook, which Kalay had earlier indicated had a strong influence in her design of assessments as this gave her the required formats and contexts she preferred to use.

I: Ok, right. And because you’re saying the textbooks give you all the kinds of aspects that you’re looking for. The different…

K: The different levels…It gives you the different formats … It gives you the contexts that are suitable Sometimes you have to, you have to apply your own creativity to, to…

In FR 1, The IF was also CSA response questions that involved representing a ratio

in symbolic form or calculating the rate of a quantity. No MCQ were given. In MR 1, CSA type of questions required learners to write the required ratio from pictorial representations. While Cramer, Post and Currier (1993) argue that these item types have "facilitators" which enable linkages between the real world and the mathematical world, there was limited link to realism in the images provided in MR 1.

Summary

In summary, Mary's tasks were contextually problematic and mathematically more limited than Kalay and Fiona. The required range of skills in the task did not adequately cover the AS and tested a more limited range of skills than Kalay and Fiona's tasks. Mary's questions did not involve applications involving (part/part) and (part/whole) as observed in Kalay and Fiona's questions. **Table 6.22** provides a summary of the teachers' assessments of Rate and Ratio.

Table 6.22: Summary of teachers' assessment of Ratio

	Teacher		
Category	Kalay	Fiona	Mary
MD	Covered both aspects of the AS. All problems involve calculation skills with number range including decimals and fractions in the context of money. Also assessed are unit rates, multiplication by integer and multiplication by a fraction, time-distance relationships and halving and doubling.	Included questions on both aspects of the AS. The skills assessed are identifying two parts of a ratio and parts in relation to the whole quantity and in the context of a recipe involving doubling, multiplying by an integer and representation	Included questions only on one part of the AS: comparing quantities of the same kind. Assessment based only on the (part: part) ratio order skills and the only prior mathematical knowledge required is counting.
DL	Moderate to difficult. In keeping with Grade 6 AS. Questions involve scaling up and scaling down of quantities and involve "graded' sub-parts. Questions are scaffolded in order to implicitly present a range of questions with varying levels of	Moderate difficult. In keeping with Grade AS. This is due to the simple number range applied within the .Grade 6 curriculum.	Easy. Questions are lower than AS range of Grade 6. All questions pitched at the same difficulty level and simple number range.

	difficulty.		
CD	Overlap between routine and complex problem solving. Applications went beyond knowledge recall and a multiplicity of skills was assessed in each sub-question	Mainly centered on knowledge applications and routine problem solving. Limited use of fractions and decimals in the questions to extend the cognitive demand of learners.	All questions are knowledge based and require the same level of thinking, which was to refer to the given information or recall the symbolic representation of ratio from class work and apply it to four picture diagrams
LD	Mainly high text problems with decimals and money contexts. Learner is required to interpret the word problems through a critical reading of the text.	High text problems given on this topic.	High text application used. Learners had to make certain language connections between terms. The skill of inference in the context weighed heavily.
IF	Closed short answer type showing calculations. The format of the question was extracted from a textbook.	Short answer questions that involved representing a ratio in symbolic form or calculating the rate of a quantity.	Closed short answer type requiring learners to write ratio in symbolic form from pictures. No calculations required.

Capacity

a) Expected skills

The second topic for analysis involved measuring the capacity of objects. Capacity refers to how much an object holds and is usually measured in litres and millilitres (e.g. the capacity of a milk bottle is 1 litre). This fell within the content domain of measurement (LO 4) in the RNCS. Within this content domain, learners were required to show competency in measuring units, instruments and formulae in a variety of contexts. There were two assessment standards on capacity listed in the RNCS (DoE, 2002b). The first involved estimation and the appropriate use of units.

> *The learner will be able to estimate, measure, record, compare and order two-dimensional shapes and three dimensional objects using S.I.[10] units with appropriate precision for:*
>
> - *Capacity using millilitres (ml) and litres (l).*

[10] S.I. units – Systeme International (d' Unites): the universally used system of scientific units)

The same AS included the measurement of mass, length and temperature. The second AS involved the appropriate use of measuring instruments and allowed for applications to be presented in real world contexts.

> *The learner will be able to use appropriate measuring instruments (with understanding of their limitations) to appropriate levels of precision including:*
>
> - *Measuring jugs to measure capacity.*

Hart (1981) argued that learners are often encouraged to use different units to measure prior to the introduction of a standard unit. Studies on measurement (Lehrer et al., 1998; Stephan, Bowers, Cobb, & Gravemeijer, 2003) have emphasised a gradual progression from measuring with non-standardised units (e.g. estimating using cups) to measuring with standard units (e.g. a measuring jug). The literature on measurement studies indicate that the expected skills on capacity questions are generally to show: 1) the number of units needed to describe the size of the measured unit and 2) the convenience of using a standard unit. These two assessment standards indicated on capacity reflect the literature expectations of competencies to be assessed within the domain of measurement.

Literature also points out traditional areas of weakness in measurement that have been observed in research studies and are important for assessment. The analyses in NAEP and the TIMSS studies have shown that learners' patterns of reasoning on measurement questions often indicate confusion with high percentages of learners unable to answering questions involving standard units and even fewer answer questions correctly when non-standard units are used. Measurement has remained a weakness among 4th and 8th grade learners on numerous assessment cycles (Smith et al, 2008).

Hart (1981) indicated that measurement is taught in most British primary schools, and an important aspect in teaching this topic, is the appreciation of the need for a standard unit (Hart, 1981). Researchers (Stephan and Clement, 2003, p.3) have argued that measurement instruction has traditionally focused on the procedures of measuring rather than the knowledge concepts underlying them and the lack of teaching fundamental principles that justify these procedures have caused many problems for learners. Researchers working on the STEM project (*Strengthening*

Tomorrow's Education in Measurement) in the United States (US) argue that extensive evidence has shown (and continues to show) that US students' grasp of spatial measurement (e.g. Length and Capacity) is poor, despite the wealth of spatial experience and knowledge they develop and use outside of school (Smith et al, 2008).

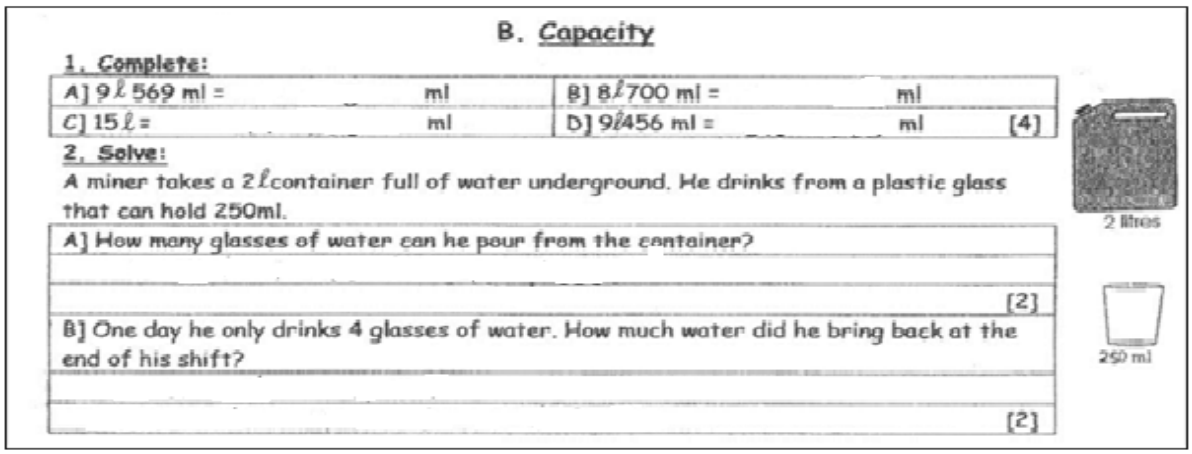
In light of these concerns, assessment tasks covering the range and scope of the second AS (listed above) are more cognitively demanding than the first AS, since there will be more measurement reasoning and calculations expected than in the first AS. Also, when contexts are used to frame the capacity item, the language demand of the second AS are likely to have more high text questions than low text ones (e.g. which glass is half full?). The expectation in national curriculum is for teachers to assess learners on both assessment standards and the literature suggests that questions should progressively range from the first AS into the second. The tasks that follow show how the three teachers selected questions and tasks on the topic of capacity in their assessments.

b) Tasks

The tasks designed by teachers on Capacity are indicated in **Table 6.23** below. In KC 1, Capacity is assessed in terms of conversion and rate. In FC 1, there is a single assessment task in which she also asked learners questions on other measuring concepts such as mass and length. In MC 1, the task was extracted from a textbook and consisted of two questions: first a question on terminology and the second question on conversion of units.

Table 6.23: Assessment tasks on Capacity

KC 1



B. Capacity

1. Complete:

A] 9ℓ 569 ml =	ml	B] 8ℓ 700 ml =	ml	
C] 15ℓ =	ml	D] 9ℓ 456 ml =	ml	[4]

2. Solve:

A miner takes a 2ℓ container full of water underground. He drinks from a plastic glass that can hold 250ml.

A] How many glasses of water can he pour from the container?

[2]

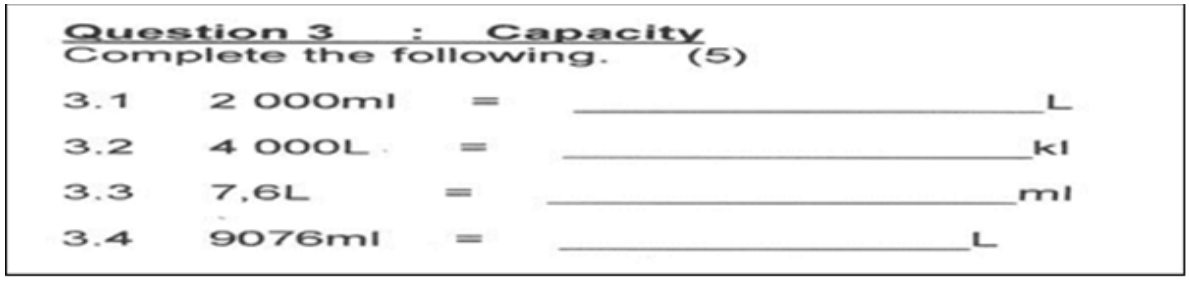
B] One day he only drinks 4 glasses of water. How much water did he bring back at the end of his shift?

[2]

2 litres

250 ml

FC 1



Question 3 : Capacity

Complete the following. (5)

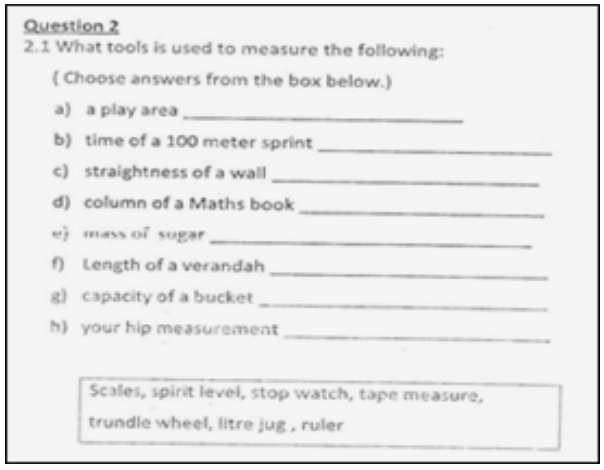
3.1 2 000ml = _____ L

3.2 4 000L = _____ kl

3.3 7,6L = _____ ml

3.4 9076ml = _____ L

MC 1



Question 2

2.1 What tools is used to measure the following:

(Choose answers from the box below.)

a) a play area _____

b) time of a 100 meter sprint _____

c) straightness of a wall _____

d) column of a Maths book _____

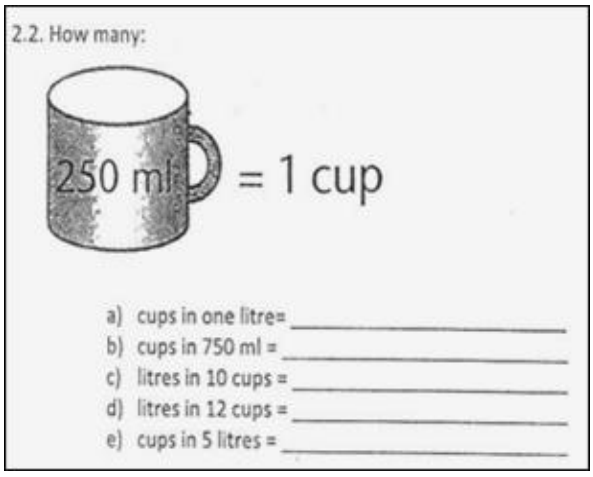
e) mass of sugar _____

f) Length of a verandah _____

g) capacity of a bucket _____

h) your hip measurement _____

Scales, spirit level, stop watch, tape measure, trundle wheel, litre jug , ruler



2.2. How many:

250 ml = 1 cup

a) cups in one litre= _____

b) cups in 750 ml = _____

c) litres in 10 cups = _____

d) litres in 12 cups = _____

e) cups in 5 litres = _____

Listed below is an analysis of how the three teachers selected questions on the topic of Capacity in their assessments.

c) Analysis

MD

All three teachers had questions on the AS dealing with the conversion between *millilitres (ml) and litres*. In KC 1, the focus was on the mechanical conversion of whole number units involving *ml* and *l*. KC 1 deals with the AS dealing with

measuring jugs to measure capacity with precision and in context which is absent in FC 1. In FC 1, only the concepts of conversion as stipulated in the first assessment standard are included, but with a broader range of measures and incorporating scaling in both directions between these measures. For example, FC 1 covers the conversion of *litres (l)* to *kilolitres (kl)*. MC 1 had two parts with the first on measuring tools terminology and the second part on both the conversion AS and the context of using measuring jugs. MC 1 was not structured with the multistep problem solving skills included in KC 1.

DL

In KC 1, the DL level ranges from moderate (in keeping with the Grade 6 curriculum) to difficult. For example, question 1 has conversions that learners in Grade 6 are expected to solve but question 2B includes the need to interpret and calculate what the "miner" drinks and brings back at the end of a "shift". In terms of the ASs for Grade 6, these types of skills are above the core grade curriculum, which increases the DL from moderate to difficult. More typical representations of this type of conversion are seen in MC 1 (question. 2.2), without the added interpretation skills, therefore reducing the DL to moderate. For example, in Mary's questions you have to work out how many 250*ml* are in 5*l*. Mary's comments differed slightly from this coding and in the interviews suggested that questions ranged from easy to moderate.

I: So you're saying that all the questions…

M: Is ok … are pitched at a kind of easy to middle or to moderate level

For Fiona, the number range in FC 1 involved both scaling up and down. There was a noticeable inclusion of decimals e.g. 7*,6l* in one sub-question that further extended the number range of the conversions. This one and last question of converting 9076 *ml* to *l,* does involve dealing with decimal values, so this can be argued to be harder in relation to these numbers not being so common in the Grade 6 curriculum. In DL terms, Fiona's items are the hardest of the three tasks.

CD

In KC 1, the tasks involved scaffolding of questions with some having multi-step procedures. For example, in question 2 the first part is directly linked to solving the

second part. The *working out the number of glasses* is routine and procedural because the operations are expected for that grade, but working out *how much the miner brought back* is more complex problem solving as there is additional thinking required in working out the remaining capacity after a part of the capacity has been consumed. In FC 1, the required thinking involved knowledge recall of converting of SI units. In these questions, there was a single step involved in the conversion. Although, the questions with decimals had higher DL levels, the CD level was still at a knowledge recall. In MC 1, the assessment of terminology in the first part was purely knowledge recall and matching. In the second task (question 2.2), the questions straddled procedural and complex problem solving as the learners had to make connections and work back and forth with the conversions of cups and litres. Where the learner has to either convert or scale up or scale up and then convert using a sequence of steps involving fractional values (e.g.$\left(2\frac{1}{2}\right)\ in\ 10\ cups$), the CD level increases from procedural to complex due to the combination of skills involved.

LD

In KC 1, the first question is mechanical with low text used. The second question is phrased as a word problem and learners have to make connections with the terms before they can generate a solution. This was a high text question with the context of the miner incorporated. As noted earlier in terms of scaffolding supports, Kalay includes a picture of a 2-litre can and a glass showing the 250 ml values in Q2. This is in contrast to MC 1 which includes a picture of the 250 ml cup, but not the litre measurement. In FC 1, the text used was almost only for computational understanding with minimal usage of language written in the instructions. This was a typical example of an instructional verb type question with low text that Fiona preferred to assess the basic knowledge skills of learners. The first part of MC 1 was high text as it dealt with terminology linked to measurement tools. Mary had commented that in questions that had a lot of context, several of her learners would struggle because of the English usage. The vast majority of learners in her classes were from communities where English was not the home language spoken.

I: And the English, doesn't present any problems for the second additional, third additional language learners?

M: It's a problem here…It's a problem, yes.

I: So sometimes it's the language that can also play a role.

M: Yes

In spite of noting this problem, Mary used of a lot of her tasks directly from textbooks and high text tasks like 2.2 were observed.

IF

For all three teachers the IF preferred by the teachers was CSA type questions with the format of questions almost similar with learners required to complete their responses in the spaces provided. Very little writing space was afforded in KC 1 and MC 1 although learners were required to perform some written calculations in their solutions. In FC 1, the tables used were similar to those she used in classwork, with learners workbooks providing confirmatory evidence. IN MC 1, the format of questions on this topic was also similar to what was in the formal district common assessment for the term.

Summary

In summary (see Table 6.20), the tasks of Kalay and Mary covered both expected assessment standards but Fiona focused only on conversions. This was striking in that while across overall SBA tasks, the range and scope of Kalay's and Fiona's tasks were similar, but observed differences were noted on Capacity. The sub-questions for all three teachers were mainly of moderate DL. However, Fiona included harder conversion tasks, moving in both directions on conversion, and including decimal values while Kalay included a high text problem pitched at the upper end of the curriculum. While Mary focused on the knowledge to routine range, Kalay and Fiona focused more on the routine to complex range. Fiona's questions were low text questions but Kalay and Mary included high text formats. The item format was closed response across all the tasks analysed. **Table 6.24** shows the comparisons on the teachers' assessments of Capacity.

Table 6.24: summary of teachers' assessment of Capacity

	Teacher		
Category	Kalay	Fiona	Mary
MD	Included questions on both of the expected ASs. Very clear focus on the mechanical conversion of units; combines *ml* and *l* in number representation. Tasks also deal with measuring capacity with precision and in context.	Included questions on one AS, dealing with the mechanical conversion of units. The number range involved both scaling up and down. There was a noticeable inclusion of decimals	Included questions on both of the expected ASs: the conversion of units and the context of using measuring jugs. Also knowledge on terminology within the broader domain of the measurement assessed.
DL	DL was moderate to difficult. Moderate items involved only conversion only. Difficult items combined knowledge elements of converting interpreting capacity units.	DL was moderate to difficult across the sub-questions. Inclusion of decimals extended the number range and DL of conversions usually seen at a grade 6 level. Fiona's items have a higher DL than Kalay and Mary.	DL was moderate with scaling up and scaling down of measurements in the sub-questions. Inclusion of fractional unknowns increased the difficulty of the question but remained within the scope of the grade.
CD	Included questions that scaffold from routine to complex problem solving. The cognitive demand is complex when Kalay includes multi-step procedures in context as well.	Mainly centered on knowledge and routine applications of computational work. some minor increase in cognitive demand of with the inclusion of decimals but applications are within the core curriculum.	Mary's range is mainly from knowledge into procedural Task 1 focused only at knowledge level, while in task 2, questions straddled procedural and complex problem solving as the learners had make connections and use a combination of skills.
LD	Low text questions used for mechanical conversions and high text questions used for solving word problems.	Only low text questions used for mechanical conversions.	High and low text questions used.
IF	Closed short answer type requiring calculations.	Short answer questions that involved only conversions.	Closed short answer type with calculations required.

Graphs

a) The expected skills

The third common topic selected for analysis was on reading and interpreting graphs. This fell within the content domain of data handling. In this topic, there were two relevant ASs listed in the RNCS (DoE, 2002b). In the first AS, the required skills were constructing graphs and interpreting data.

> *The learner will be able to draw a variety of graphs by hand/technology to display and interpret data (grouped and ungrouped) including:*
>
> - *Pictographs with a many to one correspondence and appropriate keys;*
> - *Bar graphs and double bar graphs.*

The second AS required learners to interpret graphs and show added competence in making predictions from the data represented.

> *The learner will be able to critically read and interpret data presented in a variety of ways (including own representations, representations in media-words, graphs) to draw conclusions and make predictions sensitive to the role of:-*
>
> - *Context (e.g. rural and urban, national or provincial);*
> - *Categories within data (e.g. age, gender, race);*
> - *Other human rights issues.*

The interpretation and use of graphs is an aspect of school mathematics which has appeared in school syllabi and national curriculum statements for many years (Kerslake, 1993). Research has recognised that for students to effectively utilise graphs it is not sufficient for them to just be able to directly read information from a graph. The National Council of Teachers of Mathematics (NCTM, 1980) called for an increased emphasis on drawing inferences from data (Pereira-Mendoza and Mellor, 1990). This ties-in with the second assessment standard on graphs listed in the RNCS. Regarding the general objectives of graphs, Kerslake (1993) argued that every child should know and show sufficient skills that enable him or her to appreciate the visual display of information and be able to interpret such information when it appears in newspapers and magazines. According to Kirk, Eggen and Kauchak (in Pereira-Mendoza & Mellor, 1990) the maximum potential of a graph is actualised when the reader is capable of interpreting and generalising from the data presented.

There is however an abundance of evidence through research studies on graphs (Clement, 1985; Even, 1998; Janvier, 1981; Kerslake, 1993; Sharma, 1993; Hadjidemetriou & Williams, 2003) that learners often experience difficulties and misconceptions in interpreting information represented graphically. Varied reasons are presented for this. Many learners, who are unable to treat the graph as an abstract representation of relationships, appear to interpret it as a literal picture of the underlying situation (Clement, 1985). Ainley (2000) argued that graphical work in general and interpretations skills in particular become more transparent or fused when children embed them in contextual activities. Hadjidemetriou and Williams (2003) pointed out that learning to interpret graphs in new contexts is an important but demanding skill, which requires its own practice.

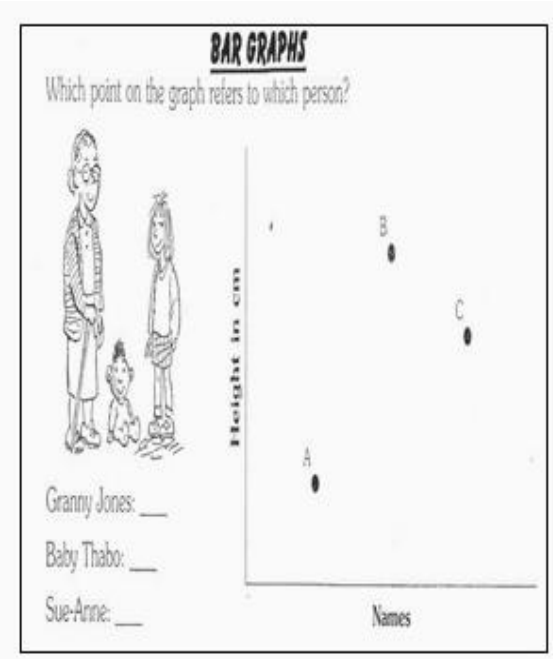
In light of these concerns, assessment tasks covering the range and scope of the second AS (listed above) would involve more complex problem solving contexts than the first, which refers more to representation of information. The added expectations of drawing conclusions and making predictions increase the cognitive demand as well. In their study on bar graphs with fourth and sixth graders Pereira-Mendoza and Mellor (1990) pointed out that while students had few difficulties with literal reading of graphs, they were often unsuccessful in answering questions requiring higher level cognitive skills. Listed below is a descriptive analysis of how the three teachers selected questions and tasks on the topic of graphs.

b) Tasks

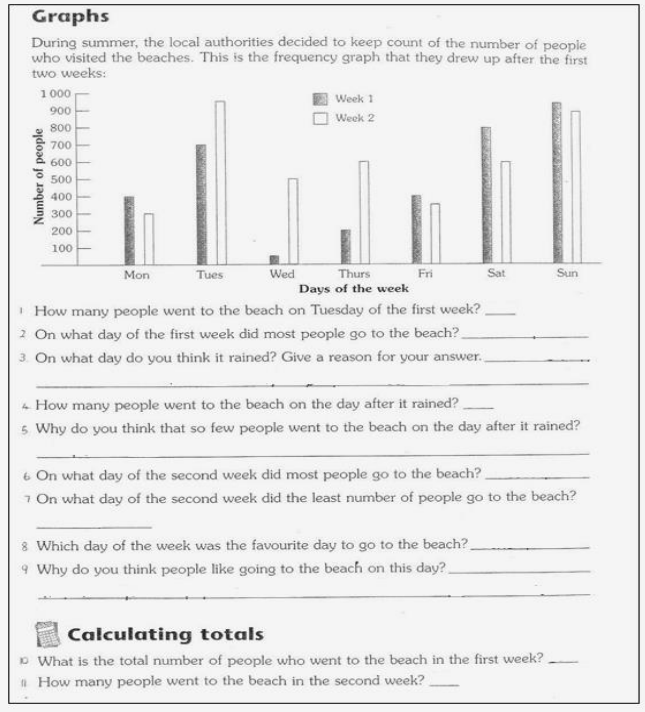
Tasks designed by teachers on Graphs are indicated in the **Table 6.25** below. In **Table 6.25**, KG 1 and KG 2 gives a sense of the varied range and scope of questions asked by Kalay. In KG 1, questions are structured on single points of a bar graph. In KG 2, a graph requiring more interpretation was given to assess skills of using a double bar graph (stacked). FG 1 and FG2 were typical examples of assessment items on graphs designed by Fiona. Questions focused on studying the graph and answering the questions that follow. FG 2 included a double bar graph with more data to interpret than FG 1. MG 1 and MG 2 were line graphs and bar graphs respectively and in both tasks, data had to be interpreted and read off.

Table 6.25: Assessment tasks on Graphs

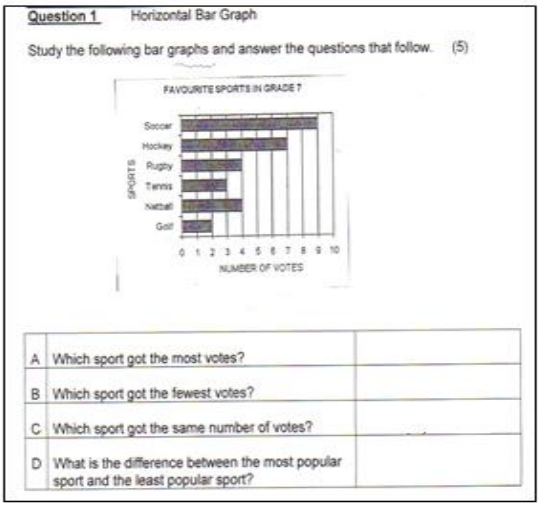
KG 1



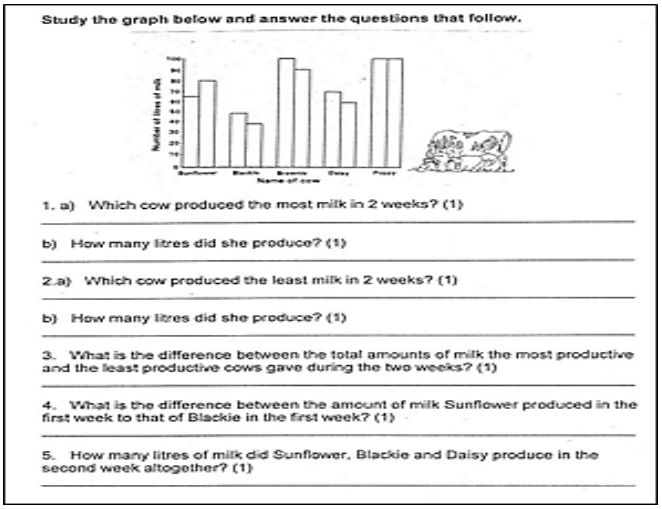
KG 2

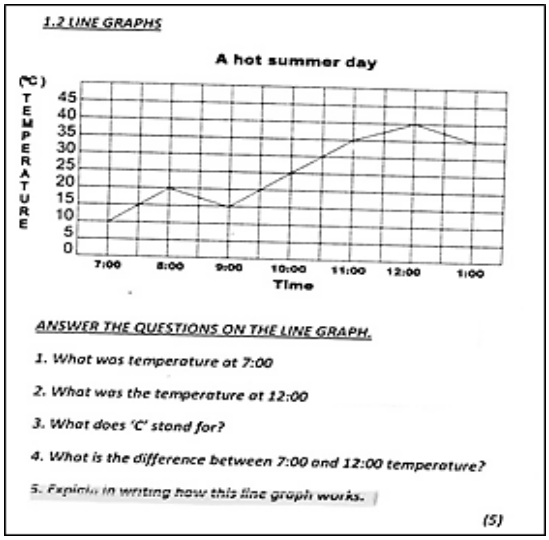
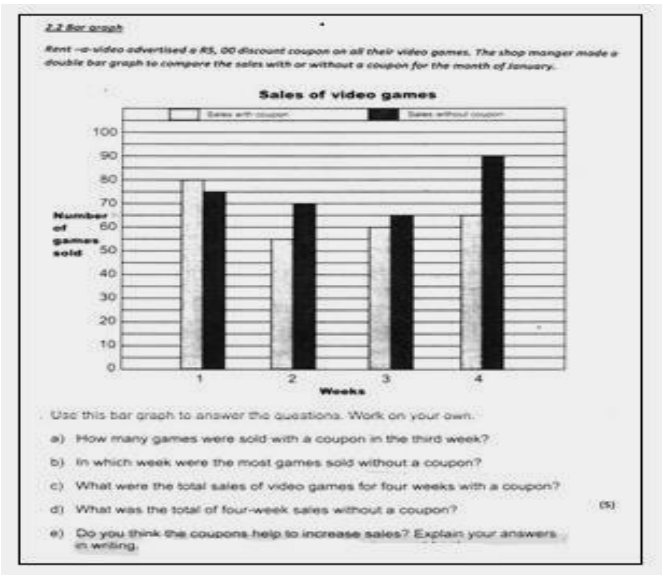


FG 1



FG 2



MG 1	MG 2
	

c) Analysis

MD

In KG 1 a basic scatter-plot graph was given to assess the skill of associating information in the text to data points and covered basic interpretation skills associated with the first AS. A lot of the information is visual. An association of person height and age is linked in the context of the question. In KG 2, a double bar graph was used to describe how a local authority was tracking the frequency of visits to a beach during different days of the week for two weeks. The representation of two sets of data had to be understood and utilised by learners. Here questions are designed to assess a learner's interpretation of the given data but there are also questions that prompt learners towards interpreting and drawing conclusions. Both ASs are covered in KG 2. In FG 1, a horizontal bar graph was presented for learners to study and then answer the four sub-questions that followed. In FG 2, a double bar graph was used to assess learners' ability to read off data and answer five sub-questions. In both tasks, an everyday real world context was linked to the data represented in the graphs but coverage was linked to first AS only and no prediction type questions were given.

MG 1 and MG 2, the tasks were derived from a particular textbook Mary was comfortable with. In MG 1, the assessment comprised of questions on line graphs. Five sub-questions were given within an everyday context on temperature and time.

A chart was given in which learners had to answer questions on the relationship between two quantities: temperature and time. Part e in both tasks required an inferred explanation based on data given.

DL

The questions in KG 1 are of the same difficulty level, which in terms of the DL framework can be categorised as easy as it involves skills from a lower grade. In KG 2, the DL was more spread out. Sub-questions were structured with different levels of difficulty. There was a noticeable build up in the complexity of skills and difficulty level. The added difficulty in KG 2 (e.g. question 9) was that an explanation was required that went beyond the data represented on the graph. As indicated earlier, Kalay was comfortable in exposing and assessing learners on high order questions. These were coded as difficult.

In FG 1 and FG 2, the DL was moderate as both tasks were in keeping with the Grade 6 curriculum requirements. FG 1 assessed learners on basic skills required in answering questions on graphs. In FG 2 there was some added difficulty to the structure of sub-questions with a double bar graph. Fiona's DL coding of graphs was slightly different. She had indicated that she chose to structure the sub-questions with generally the same difficulty level, opting more to classify different tasks on graphs as either easy or difficult rather than including a combination of sub-questions that are easy or difficult in each task. In the interview, Fiona referred to FG 1 as being fairly easy for the learners to answer and FG 2 more difficult than FG 1. .

F: If you look at the different, the first graph was a fairly easy graph, so that would have catered for my weaker learners.

I: Ok

F: Even by looking at the graph you can pick up the answers. If you looked at the double bar graph that was a little bit more difficult and a little bit more challenging for our learners. More application or interaction with the data was involved in that graph.

The DL of MG 1 and MG 2 ranged from moderate to difficult. In MG 1 questions one and two involved a basic extraction of data from the graph. In question 5 in MG 1 learners were required to provide an explanation: 'Explain in writing how this line

graph works'. This type of question was not observed in the other two topics. It was coded as difficult as learners had to infer an explanation about how the graph works – a skill pitched at the upper end of the curriculum requirements. Mary had mixed views on question 5, regarding it as general knowledge while at the same time acknowledging it was more difficult than the preceding questions.

I: Now how would you…Would you also say that question is fairly easy or is it, is it more demanding than…?

M: It's more demanding.

I: Ok, because what was expected?

M: It's general knowledge … They must think …They must think and then describe whether, how it's moving from 7, from the different time intervals.

Mary's statements did not provide clarity on what would be evaluated as a correct answer, which has been noted as a broader problem in South African research (Hoadley, 2012). In MG 2, all the questions set, apart from the last one, were coded as moderate DL, since they fell within the Grade 6 scope. The last question was coded as difficult due to its being pitched above the grade requirements. Mary though, considered all the questions as easy.

I: Overall, before we get into the questions, how do you feel this kind of question was? Was it easy or difficult?

M: Easy. It's not difficult, it's easy just because…Because at grade 6 level they do double bar graphs.

It should be noted that Mary provided additional support by means of explaining instructions orally to learners. This was one way of reducing the difficulty level of the question so that learners could cope, but there is also little specific detail in Mary's comment about what doing 'double bar graphs' might entail. Given evidence of difficulties for teachers in dealing with data handling ideas and connections coherently (see Venkat & Adler, 2012), it could also be seen as a need for some interpretation when assessment instructions were unfamiliar to learners.

CD

In KG 1, the CD was mainly knowledge recall. There was no extensive problem solving involved unless a learner could not make the link between individuals of

different ages and their corresponding heights. There were no units or scale to work with and no calculations needed to answer the question. In KG 2 the CD was spread between routine and complex problem solving. In this extract, questions were designed that required learners to think and apply knowledge beyond the data represented. The first two questions were basic extractions of data at different time points. In questions three, four and five an interpretation of the data and the context of visits to the beach was required. There had to be an understanding and knowledge that on rainy days, there are fewer visits to the beach. Questions 6 and 7 required learners to understand mathematical applications involving the terms “least” and “most”. In questions eight and nine, an interpretation of lifestyle and recreational choices was also explored and learners had to offer an explanation of why a high number of visits took place on Sunday during the first and second week. Pereira-Mendoza and Mellor (1990) had pointed out that these types of questions had a high cognitive demand.

In the FG 1, the first three questions the CD knowledge recall and the fourth was routine. Common mathematical terms such as “most”, “fewest” and “the same” were assessed. In the fourth question (D), an additional calculation had to be made based on earlier answers. In FG 2, the questions were also coded as routine. In the first two questions again deal with finding out the “most” and the “least” with the added context of “productive” attached. Also, the word “least” is used instead of “fewest” bringing in the assessment of similar terminology used in different contexts. In questions 3 and 4, working out the difference is tested. In question 5, the additional skill is for a specific addition to be made. The word “altogether” is used to trigger the required working.

The CD of MG 1 and MG 2 were routine but in each task, question 5 was more cognitively demanding than the previous questions asked and was coded as complex. In MG 2, question five was an interpretation question: ‘Do you think that the coupons helped to increase sales. Explain … your answers in writing.’ Several calculations had to be done before an appropriate answer could be given.

LD

The LD in KG 1 and KG 2 were high text as a result of the context underlying the

data in the graphs. In all questions, the understanding of the text was important to find the solutions although it did not overtly distract from the mathematics required. In KG 2 the LD of the context was significant in learners presenting viable solutions. In this type of problem solving, Kalay explored the use of open ended questions which was not always possible with other topics. She was comfortable in extending the thinking and reasoning ability levels of learners even though some would not have had experiences with visits to the beach. The LD in FG 1 and FG 2 was high text with real world contexts used to situate the data. In FG 2, the amount of text is more language intensive than the text used in FG 1 (on sport). Across both tasks the text used had basic instructions and mathematical terminology and learners would be able to understand them.

In MG 1 and MG 2 assessment tasks were high text and involved the understanding of concepts such as discount and increasing sales. It was indicated earlier that with high text problems, where there were unfamiliar terms, there would be some oral interjection to assist learners to understand some of the concepts.

I: Ok. So would you have added any information to further help the learners understand discount?

M: Yes. Explain first.

I: Oh, you would explain? But this was a test now. In a test situation would you…even in a test…you would still…explain to the learners to understand.

M: Not all, the difficult ones.

I: The difficult ones? Because a word like ‘discount’ he (the learners) would have done it…

M: Yes, If you explain before … you can at least get a better mark.

In particular, Mary gave additional explanations to learners dealing with more difficult and cognitively demanding questions such as question five in MG 1 as most of them were second language speakers. Stein et al (2000) noted that CD is often reduced in the enactment of classwork, reflected in Mary’s acknowledgement that she didn’t leave learners totally on their “own”.

M: Ai, I can’t

I: You can’t

M: I can't

I: Because you feel, you feel that the learners need that kind of additional ..?

M: Attention, yes

IF

The IF of KG 1 and KG 2 were CSA drawn from textbooks that learners also used in their classwork, so there was a sense of familiarity. The format allowed for learners to accurately read off the data. Similarly in FG 1 and FG 2, the format of items was CSA sourced from textbooks that were used in classwork activities. Questions were reformatted in the form of tables to facilitate learner responses. In MG 1 and MG 2, the item format was also sourced from textbooks but the graphical representation did not always allow learners to extract information accurately. In MG 2, it was not obvious that half way is exactly half way between two intervals (e.g. 60 and 70) which is required in answering the fourth question in MG 2. Mary penalised inaccurate answers.

I: So even with the fourth question it falls in between. Did any learners give an answer of 66? The answer is 65

M: Yes

I: Did some of them say 66 or 67?

M: But the answer is 65.

I: But what did you do in those cases when they give 66?

M: It's wrong.

More clarity on the unit scales in this type of CSA questions would have reduced some unintentional ambiguity.

Summary

The assessment tasks described above summarise Kalay's efforts to assess as fully as possible a wide range of skills at different difficulty and cognitive levels covering both required assessment standards. The assessment tasks of Fiona were adequate in terms of coverage of the ASs but did not extend learners towards more demanding application and problem solving type questions. For Mary, the tasks assessed a more varied range of skills across the ASs in comparison to Fiona and Kalay but were not as cognitively demanding. Mary's questions ranged across from

easy to difficult. **Table 6.26** below shows a comparison of the teachers' assessment tasks on Graphs.

Table 6.26: Summary of teachers' assessment on Graphs

	Teacher		
Category	Kalay	Fiona	Mary
MD	Included questions on both of the expected ASs that assessed learners' abilities to interpret data in routine and complex problem solving applications of graphs.	Included questions on one AS, dealing mainly with routine applications of interpreting data from different forms of graphs	Included questions on both of the expected ASs with learners expected to present explanations at the end of the tasks.
DL	Overall DL was moderate in keeping with grade level AS. In KG 2 the DL was more spread out than in KG 1 with sub-questions structured against different levels of difficulty. There was a noticeable build up in the complexity of skills and difficulty level.	Overall DL was moderate in keeping with grade level AS but contrary to the other two teachers DL structure, Fiona chose to structure sub-questions in tasks with generally the same difficulty level, opting more to classify different tasks on graphs as either easy or difficult	The overall DL ranged from moderate to difficult with questions assessing basic extraction of data to ones that required an explanation of the data presented on the graph. The latter were pitched at the upper end of the curriculum.
CD	In some tasks the cognitive demand was mainly knowledge recall while in others the cognitive demand was spread between routine and complex problem solving. In KG 2 questions were designed that required learners to think and apply knowledge beyond the data represented.	The CD ranged from knowledge recall to routine applications involving reading, interpreting and performing calculations from the graphical data presented.	The CD was mainly routine but in each task but there was a build up with earlier questions less cognitively demanding than the ones that followed. The last question in each task was complex requiring an explanation and working before an appropriate answer could be given.
LD	High text questions used as a result of the context of the data in the graphs. The understanding of the text was important to find the solutions although it did not overtly distract from the mathematics	High text questions in all tasks used. The text in FG 1 was more extensive in terms of language and words used than FG 2. Across both tasks the text used had basic instructions	Only high text questions are used in tasks. The LD is context dependent and learners need to be familiar with concepts such as 'discount' that is integral to the solution. For Mary, her second

	required.	and mathematical terminology.	language learners were assisted by oral interjections by herself to assist learners understand instructions.
IF	The IF was all CSA questions requiring learners to interpret data from the given graphical representations. The representations allowed for learners to accurately work with the data.	All questions were CSA type. The format of items was sourced from textbooks that were used in classwork activities. Questions were reformatted in the form of tables to facilitate learner responses.	The questions were also CSA type. The item format was also sourced from textbooks but the graphical representation in MG 2 did not always allow learners to extract information accurately.

From the above summaries, a more mixed picture emerges across the individual topic tasks of the teachers than was observed in the overall patterns on coverage and the range and scope of questions in SBA. In the earlier patterns observed on the coverage and range and scope Mary comes out as more definitely different to the other two, even with district mediation. Here, in the assessment of common topics, there are more similarities and overlaps between Mary and the other two teachers in terms of DL and CD in particular, than what was observed in the earlier analysis.

6.6 Conclusion

The analysis in the initial part of the chapter explored the rules of SBA in relation to motives and goals from documented evidence in policy texts. A discourse analysis using Morgan's frame showed that the motives and goals that influenced the design rules evident in each centre was varied but the enacted activities reflected a curriculum implementation discourse.

To further understand the complicated relationship between rules, goals and tools, AT principles were used in the study to present an EMT analysis of SBA. The EMT analysis showed that there are tensions and contradictions leading to slippages between advocated design rules in relation to goals and tools within the activity systems across school type: a suburban school (S), an inner-city school (IC) and a township school (T). It highlighted the existence of primary and secondary contradictions within and across the SBA activity systems of the three schools and

provided specific insights of how design rules are shaped by interacting elements within activity systems. The EMT analysis pointed to:

a) Kalay having to deal with multiple objects in the suburban school where contradictions exist between curriculum expectations on exposing learners to a full range of knowledge and skills and the school demands for high pass rates and throughput in their system.
b) Fiona having limited substantive professional support at the school regarding assessment design and content selection, and like Kalay the psychological goals evident in her initial designs are in conflict with curriculum implementation goals of the school.
c) The mediation of the district leading to limited contribution and ownership of assessments in the Township school, resulting in fewer tasks being designed by Mary. Critically, there was limited “backwash” effect of district designs on Mary’s tasks.

Comparatively, the suburban and inner-city school had similar motives and goals, but with varied forms of in-school support. Kalay had the greatest in-school mediation in terms of moderation of her assessment item selections. Mary’s operational conditions were deliberately influenced by district mediated goals and tools.

The second part of the analysis was structured around three themes: mathematical coverage, range and scope of questions and the assessment of common topics. On coverage it was observed that the teachers had very different SBA programmes with different emphases on selecting LOs and ASs. The coverage of mathematical content in all three schools was significantly different to the advocated requirements of the RNCS. There was also a stark contrast between all three teachers and the RNCS requirement with the teacher weightings considerably lower than expected. Mary’s SBA with the district assessment remained distinctly different to Kalay and Fiona’s SBA portfolios, suggesting that while some useful ‘bootstrapping’ was occurring through district-level support, a substantial gap remained between the T-school and the more advantaged schools.

Using specified categories of DL, CD, LD and IF to analyse the range and scope of

questions at an item level, the results showed significant variations across the three schools. Fiona's items had the highest DL but Kalay's items were more cognitively spread out. It was significant to note that district papers in T-school remained substantially adrift of policy specifications on MD balance and on CD, so this additional mediation level was not working to close the socio-economic gap. In the T-school, the textbook was the key resource and an individual assessment discourse was limited and sometimes absent with more clear design rules coming through in the Inner-city and suburban schools. Across all three schools, there was an absence of a documented design feature framework that guided the construct of summative examination style assessments.

The assessment of common topics on Ratio showed stark differences between Mary and the other two teachers, with tasks often drawn from lower grade levels, of lower CD, and variable clarity in language. On Graphs, Mary's DL was higher than the other two teachers and on Capacity, she included both high and low text questions Overall, on Capacity and Graphs, the tasks had more similarities than differences with greater overlaps among the teachers on DL, CD, LD and IF than on Ratio, indicating a much closer gap among schools on individual assessment topics than on summative assessments covering a range of topics. This again drew attention to the absence of an assessment framework for design rules.

The final focus of the study was to establish findings and tensions in relation to design rules in SNA and SBA within what was advocated, and then compares these tensions between the SBA of teachers and SNA, with the externally constructed ANA as the key artefact. These issues are the focus of the final chapter.

7. Summary, key findings and conclusion

7.1 Introduction

This chapter concludes the study. The focus here is on looking across the findings from SNA (Chapter 5) and SBA (Chapter 6), and discussing issues and implications from looking across them. I discuss the ways in which contradictions within SNA and SBA feed into, and are further exacerbated by, disjunctures between the two systems. The chapter begins with a brief noting of key findings from each of these chapters before going into the discussion across these.

7.2 Summary of key findings within SNA and SBA

7.2.1 SNA

In relation to the first research question, evidence in Chapter 5 showed that within an EMT context of multiple contested SNA goals and flawed enactment of the design rules; critical disjunctures in SNA rules were evident. Firstly, there was significant flouting of rules relating to the percentage weightings of content coverage in all three rounds of the ANA, leading to some LOs being over examined and others neglected. While, as expected, the coverage of ASs were significantly less than the total number of ASs in all 5 LOs, the effect here was that of reduced item coverage in certain ASs (e.g. interpreting information from graphs) resulting in limitations in the diagnostic and formative information that could be extracted from the ANA tests. These findings are largely consistent with SNA literature noting that national ministries of education are primarily often concerned with summative evidence providing data for comparing schools and districts' relative 'effectiveness' while calling for, but not really dealing with the requirement for evidence-based policymaking that can contribute to observable improvements in the quality of student learning (Greaney & Kellaghan, 2008; Ross & Genevois, 2006; Husén, 1987).

Secondly, on the range and scope of questions, significant differences were identified on DL and CD with again, notable deviations from advocated rules in all three ANA rounds, and in some CD categories like routine questions, by more than 20 percentage points. Almost no NT computational problems in the strictest sense could be found in the ANAs. The number of CSA in the ANA was high and different

to international assessment norms observed in TIMSS, PIRLS and SACMEQ where the majority of items are MCQ type.

7.2.2 SBA

The empirical data in Chapter 6 showed noteworthy findings on the SBA design feature rules in relation to the first and second research questions of the study. In relation to the first research question, the key findings showed several slippages between enacted assessment tools of schools and the advocated rules for SBA. One of them was the significant slippage in the coverage of assessment standards, which was found to be less than 60% of the RNCS requirement across school type. The EMTs located the slippages within a broader SBA activity system context and confirmed two distinct stories around the slippages. It was found that contested motives and goals leading from multiple assessment discourses underlie the design rules of teachers with the take-up of rules in the township school distinctly different to the suburban and inner-city school.

The motives and goals in the suburban and inner-city schools were similar; there were within-school pressures on following design rules relating to throughput (linked at some levels to psychological discourses, but with the need for passing for promotion to the next grade more predominant). Critically, this pointed to the assessment discourse in these schools shifting not as a result of cognition (psychological development) but as a result of throughput and resistance to learners repeating a grade. This was aligned with the literature on the likelihood of tensions within SBA (de Lange, 1999; Van der Berg, 2011; Darling-Hammond, 1989).

In relation to the second research question on variation across school settings, the slippage was more striking in the township school, leading to a reduced application of rules with district level mediation tools unable to close this gap. Mary's case was different to the other teachers and the analysis showed up distinct influences that were internally contradictory. With more limited specification of motives and goals the analysis in the township school showed that the textbook largely stands in for the curriculum, with no direct reference to curriculum evident. Even though there were district based assessments available at the end of each term there was, firstly,

slippage in the range and scope of items from the district in comparison to the national design rules guidance, and secondly, no deliberate attempt by Mary to align her internal tasks to that of the district based assessment. This pointed to a preference for working with intermediary texts rather than state level curriculum texts. Also, there was little evidence of any aspects of the psychological discourse. Differences between the township school and the other two was also evident in the assessment of common topics on Ratio, with tasks often drawn from lower grade levels, of lower CD, and variable clarity in language.

7.3 Looking across SNA and SBA: discussion of findings and implications

When looking across SNA and SBA, which was the focus of the third research question of the study, the discussion on findings is articulated on the two levels that have been used in this study. The first level relates to an EMT overview analysis, taking in motives and goals that underlie a complicated relationship between advocated design rules and enacted tools. The second level considers a mathematical analysis of tools by comparing coverage and range and scope of questions with what was advocated. Findings are discussed and then unpacked in terms of implications and relevance for policy makers, district subject specialists, school-based practitioners and researchers.

7.3.1 Level 1: EMT overview: motives/goals/rules/tools

The following observations were evident:

1) The first observation emanating from the analysis was that motives and goals from the EMT of SNA did not always influence SBA motives and goals. This contradicts the advocacy in SNA texts, evidenced in South African national and district texts, that assessment motives and goals are structured for a *nested* take-up of design rules (Jaworski and Potari, 2009; Nunez, 2009). However, the evidence indicates that teachers are often coerced by national ministries of education to adopt state driven policy goals and teach in ways that are consistent with curriculum implementation objectives (Morgan, 2000).

The analysis of texts in this study showed that there were contested goals arising out of policy and school based systems, making the nested take-up problematic and

unrealistic. Instead, at the level of implementation, school texts showed that enacted tools indicated more of an *interacting* assessment activity system where motives and goals stemmed not only from a hierarchical internalisation of the national policy driven system but also from a school-based system with pressures on throughput and preferences for working with intermediary texts. The evidence in this study suggests the latter pre-dominate the SBA design feature context (see **Figure 7.1)**

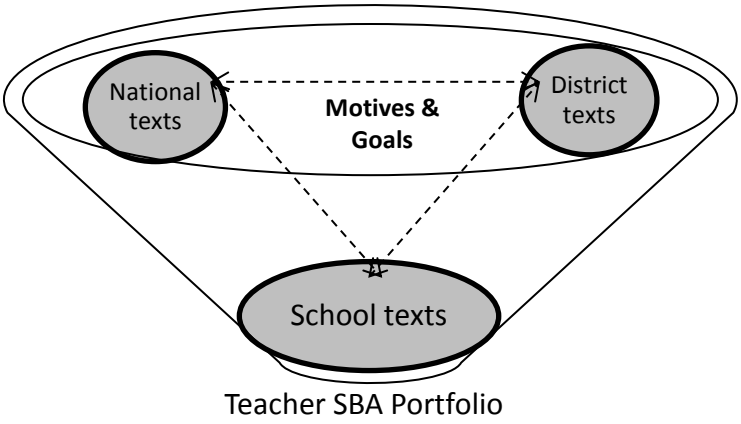


Figure 7.1: Elements of interacting policy layering

This finding, which emerged quite early in the analysis led to the decision to formulate SNA and SBA in terms of interacting, rather than nested, assessment activity systems, as there was too much evidence of influence on SBA coming from beyond the SNA system directly. Interacting rather than nested systems also suggested that there were some school-level goals that did not always coincide with policy-level goals. At the design level, there was evidence of ongoing interaction between the school and district levels and national assessment tools, but national and district texts did not have a solid linear cascade or "con-cyclic" embedded effect on the SBA practice of teachers, as Nunez (2009) found in his work relating to the mathematics educational community more broadly.

2) A critical consequence of the disjuncture in SNA, is the lack of a substantial "backwash" effect of design rules, which state ministries of education depend heavily on as reported in the literature on high-stakes assessment (Anderson and Morgan, 2008), from assessment artefacts (e.g. the ANA) from the SNA system into the SBA system.

Engeström (1997) referred to this type of contradiction, occurring within a network of

activity systems, as quaternary contradictions. At the level of advocated rules, this study showed that quaternary contradictions existed between SNA and SBA rules because they were based on different assessment discourses but linked to a joint outcome. SNA rules were, in the main, based on a curriculum standards assessment discourse but at the SBA level, the discourses were more aligned to curriculum implementation levels with more occasional psychological considerations. Therefore, at certain points of implementation in classroom practice the different assessment discourses are in conflict with each other. In the ANA tests, curriculum standards took precedence over psychological considerations for the learner but in the SBA tasks of teachers, a curriculum implementation and psychological discourse were more favoured, contributing to a limited "backwash" effect of SNA design rules into SBA rules.

The available evidence on South Africa's participation in SNA studies such as TIMSS and PIRLS, and in the ANA have shown that the quality of learning outcomes is very poor, implying the SNA rules are not having the desired effect. A further critical point noted in the last chapter was that SA-SNA policies' advocated rules for design are largely, neither detailed nor explained as key 'design features' in policy communication (teachers in this study showed relatively vague understandings of several of the design features). This restricts take up and use of these design features in SBA, thus further limiting the intended "backwash" effect of the ANA on SBA.

<u>At level 1 (theoretical implications):</u>

In terms of implications related to these findings, it is hard to see a way out of having multiple goals as all of the goals identified in the SA-SNA context continue to be seen as necessary. The study adds to the AT research community who look to contradiction identification within and across interacting activity systems for an increased understanding of them. Of interest in the South African assessment context is that while policy documentation and tiered policy layering suggest that the SNA and SBA systems operate within nested levels with SNA motives and goals defining SBA goals at school level, this study's evidence on design features necessitated a rejection of the nested view of interaction (Nunez, 2009), as evidence

pointed towards other significant contributing factors on SBA design other than SNA rules. Notably the textbook was a preferred source of possible assessment tasks over intermediary texts such as district common examination papers, with limited reference to SNA design features in the selection of tasks. The analysis of disjunctures between advocated rules and enacted tools in SNA and SBA activity systems suggested instead, the need to look at the two systems as interacting rather than nested.

By doing so, there are valuable insights for researchers on policy studies to extract from using comparative EMTs. Engeström (1993) argued that a unit of analysis between systems and structures on the one hand and daily classroom practices on the other is the *middle level* of tensions between the "formal structure of school systems and the content and methods of teaching" (p. 76). In this study, the middle level of disjunctures was spread primarily across motives/goals, tools and rules. On motives/goals, the discourse analysis using Morgan's (2000) frame showed multiple discourses present with contested motives and goals, significantly contributing to the non-alignment of SNA and SBA. There was evidence of broader motives from the education system relating to throughput rates to the next grade and inclusion at play in SBA. On rules, the general lack of explicit and concrete reference to SNA design features was stark, and exacerbated by differential distribution (discussed in more detail in the next section) across SES. These disjunctures fed into disjunctures at the level of assessment tools between SNA and SBA (see **Figure 7.2**).

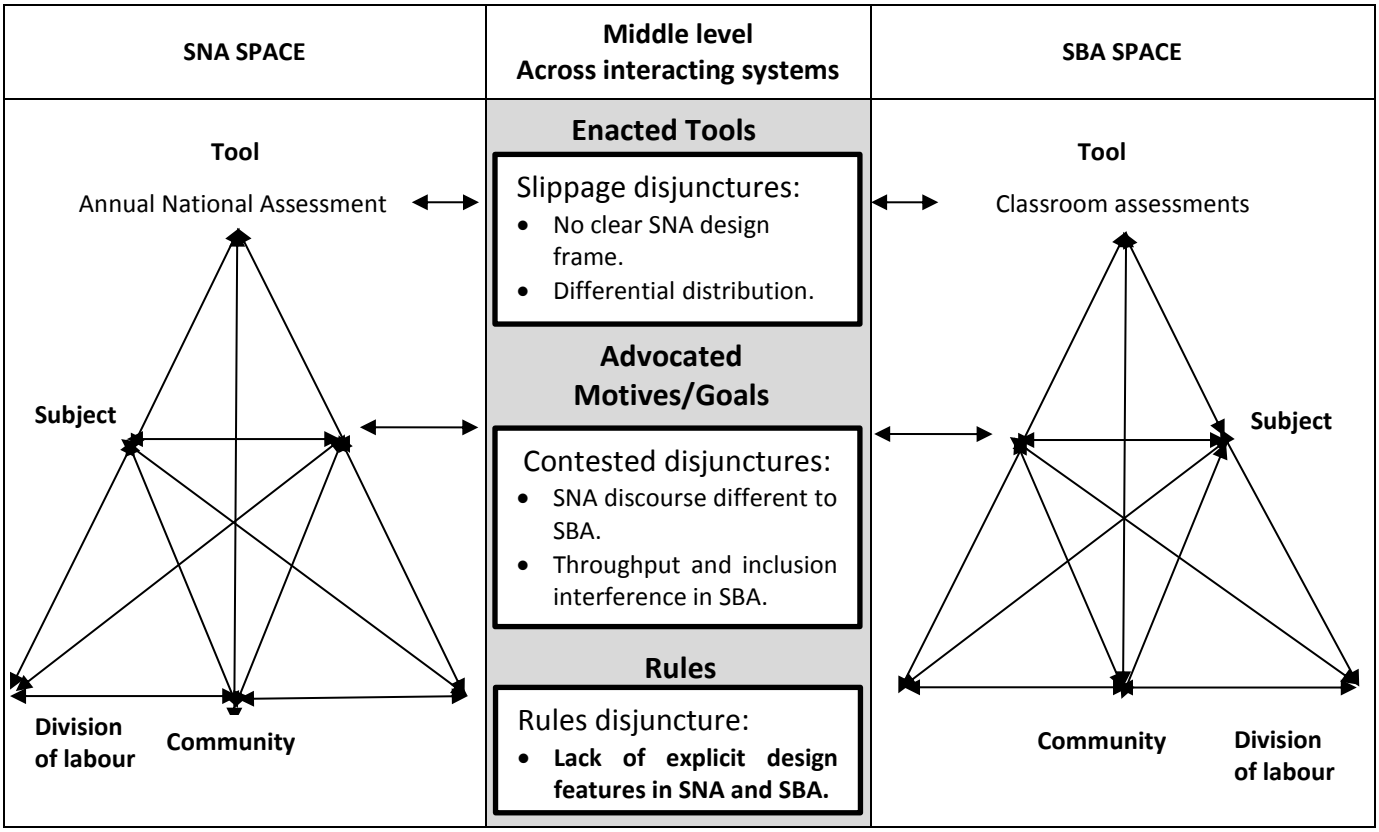


Figure 7.2: Parallel spaces across interacting activity systems

The EMTs also show comparative differences among school type with policy driven system, which in this study is important for policy makers and district officials to understand. The EMTs confirmed two distinct stories linked to school setting and justified a position to consider teachers' design rules in different socio-economic status (SES) settings. Different socio-economic status (SES) has been flagged in the SNA literature with studies such as TIMSS and SACMEQ showing SES to be a significant factor in the South African context and influencing the assessment results of learners (Spaull, 2015).

The empirical evidence on teachers' SBA in this study further supports the position that historically different school types in South Africa continue to have different access to resources and professional support (Jansen, 2009; Spaull, 2008; Taylor and Vinjevold, 1999), with district level support, as seen in this study, seemingly unable to close this gap through the provision of assessments that are more closely linked to the advocated design rules. In the township school EMT, district level mediation did not provide tools that helped to close the gap on implementing

advocated rules, and the gap between SBA tools and advocated rules was greater in the township school than observed in the inner-city and suburban schools, even though in the latter schools, district tools were not prominent. Further, comparative EMTs in different SES showed how the three Grade 6 Mathematics teachers differed in their take-up of the advocated design features, while also pointing to ongoing differences in levels of support and access to resources. The outcome appears to be different emergent norms in the application of rules in actual artefacts. This finding confirms the need to look at teachers' assessment tools both as individual and contextual constructs that are constrained by individual knowledge and access to resources.

The evidence thus suggests that instead of a mutually constituted shared outcome of assessing learners' capabilities in mathematics, SNA and SBA are often constituted instead in almost parallel spaces across interacting systems.

7.3.2 Level 2: Mathematical analysis of assessment artefacts in SNA and SBA:

A critical part of this study was to investigate whether historical discrepancies in mathematics test results between artefacts of SNA and SBA was due to a non-alignment of rules between the design of internal SBA and the design of SNA. This study shows that there is a critical non-alignment between the enacted SNA tools and the enacted internal SBA of teachers, seen from evidence contained in actual mathematics assessment artefacts (e.g. end of year school examinations). The empirical data analysed against mathematics design features drawn from a wide literature base discussed in Chapter 2 (see Table 2.14) showed that SNA and SBA artefacts overlap and contrast with the rules identified in the SNA context from an analysis of national policy documentation and the ANA assessments in the SNA context, viewed as SNA artefacts.

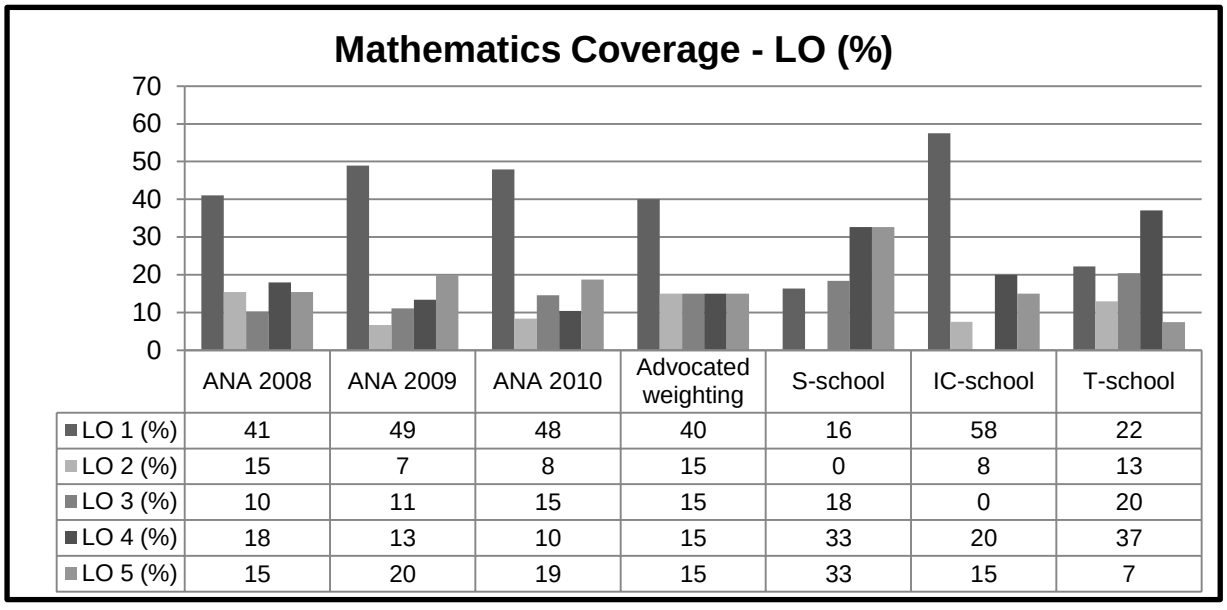
This study showed that the design features of 1) coverage across the content of mathematics, 2) level of difficulty in questions, 3) cognitive demand, 4) language demand and 5) the item format are useful item characteristics to apply when looking at specific assessment tools across SNA and SBA. Further, techniques have been illustrated on how these item characteristics can be summarised into item maps and compared. SNA literature shows similar design features are quite prominent in

established cross-national and regional SNA programmes (e.g. TIMSS and SACMEQ) (de Lange, 1999). When looking across SNA and SBA, the empirical data in these categories provided a more nuanced look at slippages in the rules governing assessment design of SNA and SBA.

The following observations are noteworthy:

1) There is a lack of alignment in content coverage between enacted tools for SNA and SBA.

Significant slippages on content coverage were observed in the differentials with expected weightings of SNA. For example, against an expected weighting of 40% (DoE, 2003a), the coverage of LO 1 ranged from 41% in 2008 to 49% in 2009 (see **Figure 7.3**). This contrasted quite sharply to coverage of teachers in the formal examinations designed teachers in their SBA, which ranged from 16% to 22%. In the IC-school the percentage coverage of LO 1 (58%) was not only significant higher than Kalay (S-school) and Mary (T-school) but also higher than any of the three ANAs from 2008–2010. **Figure 7.3** shows the stark disparities in coverage of LO among the schools themselves and the ANA to the advocated weightings of the RNCS.



	ANA 2008	ANA 2009	ANA 2010	Advocated weighting	S-school	IC-school	T-school
LO 1 (%)	41	49	48	40	16	58	22
LO 2 (%)	15	7	8	15	0	8	13
LO 3 (%)	10	11	15	15	18	0	20
LO 4 (%)	18	13	10	15	33	20	37
LO 5 (%)	15	20	19	15	33	15	7

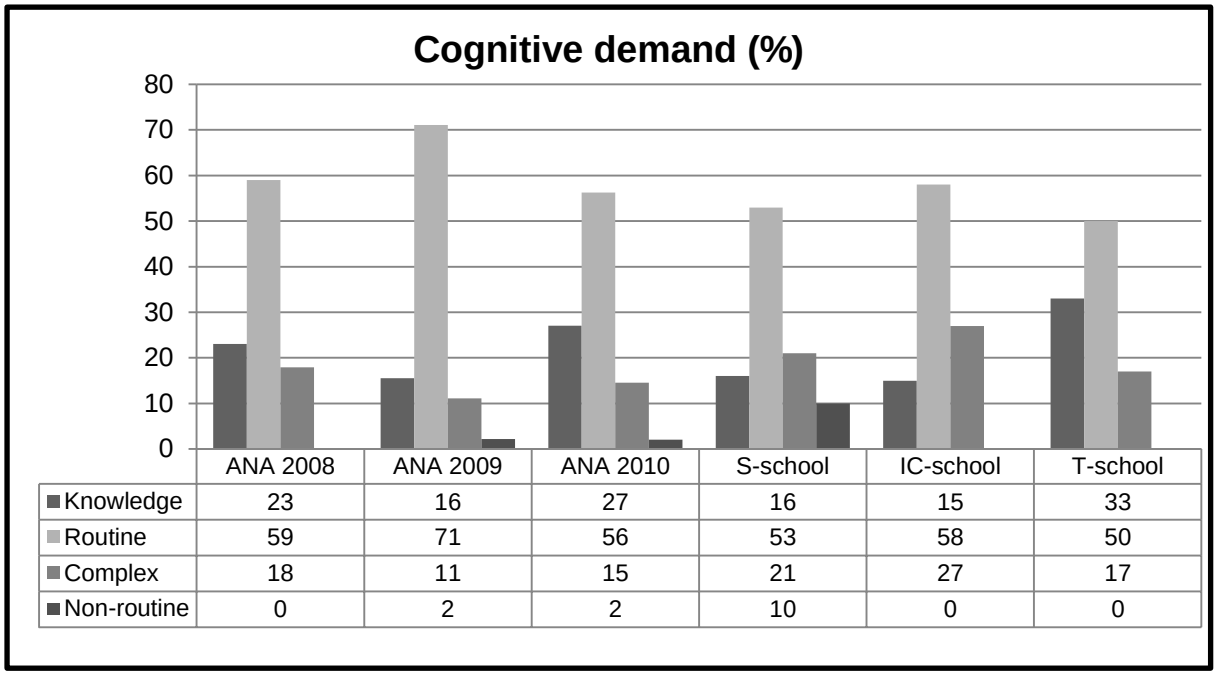
Figure 7.3: Mathematics coverage: SNA and SBA

Overall then, two results follow: firstly, there is limited match in all three school settings between the weightings across LOs to any of the ANA tests, indicating non-alignment of tools. Secondly, there is broad evidence that the SNA rules through

weightings on curriculum emphases are not being taken up in either in the ANA or across schools with different settings, with the district assessment in the T-school also deviating.

2) There were significant deviations in range and scope of questions between enacted tools for SNA and SBA.

Across the three ANAs, more than 70% of questions fell into moderate DL category and across the teachers' formal examinations this percentage was reduced ranging from 56% for the township to 60% for the suburban school. The suburban school had the widest spread of cognitive demand questions with the highest number of non-routine items, even higher than that of the three ANAs. The district based assessment of Mary (T-school) showed the highest concentration of knowledge questions (33%) compared to that of the other teachers and the ANA (see **Figure 7.4**).

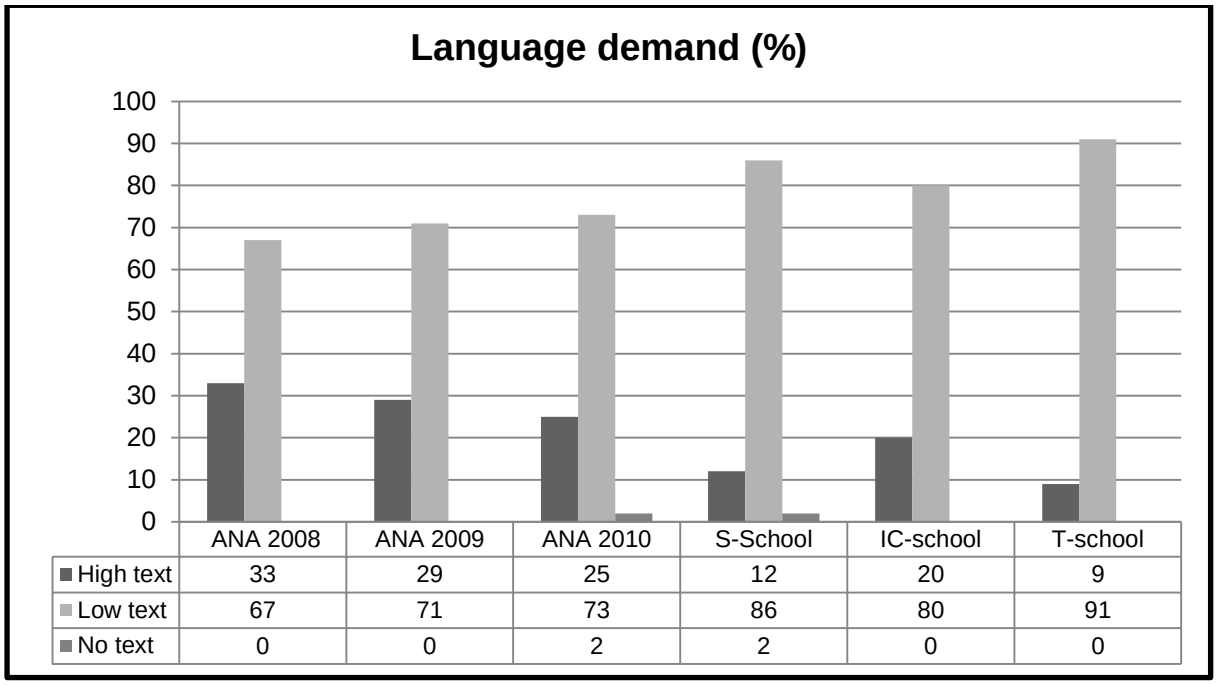


	ANA 2008	ANA 2009	ANA 2010	S-school	IC-school	T-school
Knowledge	23	16	27	16	15	33
Routine	59	71	56	53	58	50
Complex	18	11	15	21	27	17
Non-routine	0	2	2	10	0	0

Figure 7.4: Cognitive demand comparison - ANA vs School Examination

In terms of language demand, there were significantly higher percentages of low text questions (more than 85%) in the teachers' formal examinations compared to the percentage of low text questions that featured in the ANA which ranged from 67% in 2008 to 73% in 2010 (see **Figure 7.5)**. This pointed to reluctance on the part of teachers to include high text (or word problems) that historically their learners did not

do well on.



	ANA 2008	ANA 2009	ANA 2010	S-School	IC-school	T-school
High text	33	29	25	12	20	9
Low text	67	71	73	86	80	91
No text	0	0	2	2	0	0

Figure 7.5: Language demand comparison - ANA vs School Examination

Even in the ANAs, there was a reduction in the number of high text questions dropping from 33% in 2008 to 25% in 2010. Across both contexts, the inclusion of no instructional text questions was insignificant which pointed to test developers of ANA and the teachers opting to always include a stimulus in the instructions when they assessed formally.

<u>At level 2 (policy implications):</u>

An important 'take forward' point for policy makers from this study is that on both coverage and range and scope of questions there were slippages between SNA and SBA artifacts, of varying degrees across school type, with the township school the most disadvantaged. To compound matters, there was limited evidence of teachers using standardised rules from SNA to guide the design or selection of items in their assessments, leading to deviations on coverage and range and scope. Therefore, a potentially useful policy-level implication from the mathematics analysis of artefacts is for policy documents and other dissemination channels to include explicit guidance on the design features and sub-categories underlying SNA for the purposes of supporting improved SNA and SBA test construction and bridging existing gaps to SNA design rules. The vagueness of teacher comments in the interviews suggests

that exemplifications of range and scope levels in the design of mathematics tasks would also be useful.

Drawing on the empirical data analysed in this study that shows slippages in both SNA and SBA in relation to the policy-advocated design feature rules on range and scope, it would appear that further discussions and exemplifications of mapping of items to specific design features is needed within the SNA and SBA communities. A potentially useful option is to illustrate how shifts in the construction of test items play through into shifts in the mapping to design features. Discussions on this kind of work can also assist the SNA community to communicate and exemplify this framework to the SBA community, in particular, teachers, in ways that might feed in more productively into school-based constructions of assessments that better align with the policy advocated design rules.

The exemplification could be structured in the manner indicated in the **Table 7.1** below, where I begin this process by taking an example of a mathematics test item extracted from the 2010 ANA test. The original item was classified as a routine (R) cognitive demand item. The information in **Table 7.1** shows how changes in the structure of the item can exemplify a range of design feature characteristics.

Table 7.1: Exemplification of item characteristics

Nobese has 3 black, 4 red, 2 blue and 3 green balls in a bag.  Nobese takes out a ball from the bag, without looking into the bag. The chance that she takes out a red ball is _____	Original Item Characteristics CD: Routine DL: Moderate LD: High Text IF: Closed Short Answer	Description According to the item map, this item is routine because learners have to carry out or select routine operations or procedures expected for that grade. It involves a basic application of a skill or concept, in this case listing the ratio of two known quantities. The difficulty level is moderate because it is grade appropriate. There is high text usage and the item format requires a closed short answer. The answer may not be given in its simplest form.

Nobese takes out a ball from the bag, without looking into the bag. There are ___ red balls in the bag, out of a total of ___ balls altogether. The chance that she takes out a red ball is _____ out of _____.	<u>New Item Characteristics</u> CD: Knowledge DL: Easy LD: High Text IF: Closed Short Answer	If the item is restructured to have leading question parts, the cognitive demand is reduced and the combination of knowledge elements is fewer. This leading of information would be suitable for learners not familiar with ratio concepts, and below the required grade. The LD remains high and it is still a CSA type
The first ball Nobese picks out of the bag is not red. She does not put it back in the bag. What is the chance of the second ball that she pulls out of the bag being red?	<u>New Item Characteristics</u> CD: Complex DL: Moderate LD: High Text IF: Closed Short Answer	The cognitive demand has been increased due to more knowledge elements being added to the structure. The knowledge elements are still within the current grade level as learners are expected to work with ratio and rate content. However, the calculations are more advanced than the demand of original item. The DL is moderate as the actual calculations are grade appropriate. The LD is more demanding as varied mathematics concepts have to be understood. The IF is still closed.
Circle the letter of the correct answer. Nobese's chances of pulling out two red balls from the bag are: a. 1 b. Between 0 and $\frac{1}{3}$ c. Between $\frac{1}{3}$ and 1 d. 0	<u>New Item Characteristics</u> CD: Non-routine DL: Moderate LD: High Text IF: Closed Short Answer	The cognitive demand of a question can become non-routine when learners are asked to extend their thinking and mathematise situations (recognize and extract the mathematics embedded in the situation and use mathematics to solve the problem). The DL is moderate as learners at this grade level are expected to predict the likelihood of events based on observation. The LD is reduced from the original item and the IF is now a multiple choice question with 4 distractors and a selection has to be made.

Policy makers can use similar tables to assist teachers, school managers and district officials involved in mathematics test design, to map test items to a specific design feature framework. For example DL could usefully exemplify a three-level range of grade-related items (as shown in Chapters 5 and 6). A further (and simpler) implication from the analyses presented in this study relates to building awareness of the design features and their descriptors. A summary of the descriptors drawn from the literature and used in the mathematical analysis is indicated in **Table 7.2**.

Table 7.2: Design features for establishing the range and scope of questions

Design Feature	Descriptors
Difficulty level (DL)	Easy (E) (lower than grade level), moderate (M) (grade appropriate) and difficult (D) (higher than expected grade).
Cognitive demand (CD)	Knowing basic facts (K), Applying routine procedures (R), Using complex procedures (C) and Solving non-routine problems (N).
Language demand (LD)	High text (HT) (word problems), Low text (LT) (instruction/s is/are short mathematical verb/s), No instructional text (NT) (Computational).
Item format (IF)	Multiple choice questions (MCQ), Closed short answer response questions (CSA).

Current literature on SNA studies (e.g. TIMSS and SACMEQ) deals with these features in fragmented ways that admittedly are not designed for use by teachers in their SBA. For example establishing difficulty level in these studies is related to statistical procedures on item response theory which are usually only understood by a minority of teachers. Instead, taking pointers from literature more suited to SBA, such as Stein *et al* (2000) on operationalising tasks for assessing mathematics, this study suggests that policy makers and district officials supporting teachers at schools, take note and disseminate a design feature rules framework on DL, CD, LD, and IF (see **Table 7.1**) that is more relatable to for both SNA test developers and teachers when designing test items and formulating assessment tasks for classroom use. Policy makers and school practitioners should note that at the SBA level, although there were school communities in the case of Kalay and Fiona that looked at the pragmatics of whether most learners would cope with the questions, this discussion was not premised on the need for design of an appropriate test through a clear and pre-determined framework, thus representing a gap across SNA and SBA documentation.

The added value of clearly defined test specifications will be to close identified gaps between under-resourced schools, like township schools, to test design levels observed in more well-resourced suburban and inner-city schools. Although the

township school received district supplied the tools, the test framework specifications were not supplied. Mary's interview responses suggested an uncritical acceptance of the external supply, without explicit attention to aligning her own assessments with their range, scope or format, or that seen in the ANA papers. Thus, the district common papers' benchmarks were accepted without explicit consideration to aspects of their design features, and these assessments replaced the formal examinations internally set and designed by teachers. Given this, the lack of alignment of between district-enacted and nationally advocated design features, and the lack of explicit access and exemplification of these design features is particularly problematic, in that this appears to be the key mechanism for offering support to low SES schools. For district officials supporting township schools, this finding indicates a critical need to get frameworks with explication of design features to schools, and the same should be done for ANA test developers as well.

The ANA test frameworks also do not have explicit descriptors to explain their test item categories, which Linn and Miller (2005) considered an important step in planning classroom tests and assessments. Policy makers should be aware that the intended use of the ANA as an appropriate benchmark for SBA design did not appear to unduly influence the test design and item selection of teachers. Rather teachers relied on their subjective judgments based on historical knowledge of learner performance and the suitable textbooks. With the exemplification of design features through explicit descriptors, teachers can get a better sense of what the intended range and scope of questions could look like. This study illustrates how individual test items, through an analysis of the range and scope of questions, can be analysed and tagged with specific DL, CD, LD and IF characteristics. The study further showed that the design features can be used on different assessment forms from the ANAs, to formal school examinations to class based assessments on specific topics. This suggests that an exemplified framework of the design features might be useful to teachers and be useful as a tool for strengthening district capacity to support schools in selecting test items.

Within my professional role in the national assessment office, a broader scale possibility is that the design features used in this study can be used to collaboratively develop a national item bank of questions which can be used for diagnostic and

summative use. The setting up of item banks that show what the range could look like could also be set up on electronic platforms that teachers access (e.g. the Thutong portal on the DBE website in South Africa) or they could feature in the ANA diagnostic reports and disseminated to all schools. Item banks with specific characteristics based on the design features will also assist test designers of the ANA to design tests that are more aligned with the advocated rules, more than what was observed in this study.

Critically, when teachers utilise a design framework that draws on SNA rules that they feel comfortable to use in their everyday practice, a greater alignment of enacted tools within advocated rules is expected, and in the SA-SNA context a greater alignment of SNA artefacts like the ANA with teacher's SBA. This would minimise the slippages observed in this study.

The above discussion leads to my concluding commentary on the study with reflections on the process and limitations inherent within this study, and further questions raised by the findings.

7.4 Concluding remarks

To finish the moment, to find the journey's end in every step of the road, to live the greatest number of good hours, is wisdom (Ralf Waldo Emerson, 1850)

The greatest part of undertaking this academic journey was the opportunity to understand the challenges of assessment design and influences at a primary school level and make a meaningful contribution in an area of research that is limited in South Africa. The starting point of the study was to act on an assessment dilemma that has been widely accepted by researchers and the DBE that needed further investigation. The initial premise of this study was based on an assessment contradiction: South African learners continue to perform at unacceptable levels in standardised national assessments yet are seen to be making acceptable progress in school based assessments in similar age and grade cohorts. This study interprets this assessment contradiction and offers an explanation in terms of comparative design features eliciting noteworthy differences between espoused design features (analysed from rules) and the enacted design features (analysed within tools)

involved in SNA and SBA activity systems.

Within the context system tensions within and across SNA and SBA, four central ideas come through the evidence presented in this study that explains the assessment contradiction. These are:

a) The relationship between SNA and SBA is interacting than nested. At the level of advocated motives and goals, SNA texts are structured for a nested take-up of design rules but at the level of implementation, SBA design rules are shaped by not only by an interaction with policy texts but also with other non SNA rules such as textbooks, and these are sometimes preferred.

b) Multiple assessment discourses present in espoused design rules lead to contested motives and goals and these show up as disjunctures in enacted tools within what is advocated. In three rounds of the ANA, the advocated rules on content coverage and on the range and scope of questions were not adhered to. In SBA, contested motives and goals leading from multiple assessment discourses underlie the design rules of teachers with the take-up of design rules in the township school (low SES) distinctly more limited than the inner-city and suburban school (high SES).

c) Disjunctures in SBA enacted tools varied across school SES and can be linked to varied community and professional support received by teachers. Compared to the suburban and inner-city schools, there is limited professional support in the township school, leading to a reduced application of rules and the district level mediation of common test tools is insufficient to help to close the gap.

d) The absence of a common design feature rules framework, leads to a non-alignment of design rules in SNA and SBA at the level of assessment artefacts. In terms of the range and scope of test items, the ANAs are different in relation to difficulty level, cognitive demand, language demand and item format to the SBA tasks of teachers. Even, intermediary texts such as district common tests do not assist when they operate outside of a well design framework that teachers can apply in their SBA.

These ideas were explained as level 1 and level 2 findings when looking across SNA and SBA.

It is important to highlight in the South African context with historically disparate resource provisioning that comparative cases of suburban, inner-city and township schools explored in this study, involving Grade-6 Mathematics teachers, showed significant disjunctures across school SES, and the gap between SNA and SBA was most pronounced in the low SES school. The suburban and inner-city school showed a higher take-up of design rules although the main influence was learner throughput rates rather than cognition. The EMT analysis confirmed that the design rules for teachers' assessments are mediated by various interacting elements within an activity system and must be looked at broader than just individual constructs but operating within a collective Grade 6 Mathematics community constrained by individual teacher knowledge and access to resources.

Fuhrman (1999) suggested that teachers' judgements about students' performance are influenced more by their preconceptions about individual traits of students and are typically uninformed by systematic knowledge of what these students might be capable of learning under different conditions of learning. Through an in-depth qualitative study, the purpose of this research was to understand the level of interaction between the assessment contexts of SNA and SBA. This invariably implied an investigation into the nature of assessment activities that operated within each context. An appropriate way to do this was a consideration of each context as an activity system and using modern activity theory concepts to understand their structure and nature of existence. Within identified elements I could delve deeper into existing inter-relationships and contradictions that were driving the activity systems from a reference point of design rules.

The study provides valuable insights and tools that can be used for teacher training on assessment. I had earlier made the point that the relatability of a case study is of more value than its generalisabilty (Bassey, 1999). The value of this study will be in its relatability to other teachers seeking to further understand their SBA practice and for test designers at various levels of the system, to further explicate their standardised framework in ways that they can become useful benchmarks as rules for SBA. Relatability in the context of this study would also refer to researchers and teachers seeking an expanded understanding of the nature and existence of

contradictions in the dual assessment activities within the contexts of SNA and SBA. Since the study occurred among different mathematics centres, each activity system at each centre needed to be looked at both as an individual system and as interacting sub-systems part of a broader mathematics assessment system.

As the study proceeded there were opportunities, challenges and limitations that arose through the research process. This study provided evidence on a certain aspect of assessment practice; the features and influences of test design rules. The study did not intend to be an exhaustive investigation of item writing or test design process, but an implicit focus on SNA and teacher influences within their SBA. Within the limitations of an exploratory study, findings were made, but they do reflect patterns of performance, and are worth exploring further.

Undoubtedly, there are many more aspects to explore assessment practice that require in-depth investigations to fully understand why and how assessment is an area of concern in South Africa. Since the study was conducted, the curriculum has been revised and thus current findings need to take this into account. The impact of district support was not the primary focus of the study, and findings noted in this study could add to further investigation into this area. The broad intention of the study was to further contribute to existing research both locally and internationally on mathematics assessment concerns broadly and how they can be better understood by those engaging in its practice. A more narrow intention was to clarify specific disjunctures that exist in SNA and SBA and to offer an informed perspective on why they exist, as a way of better understanding the assessment dilemma in South Africa. To this end a great number of good hours were wisely spent. Immediate further work would be to extend the analysis of ANA papers in the 2008-10 cycles in this study to the papers of recent years, to see if alignment to advocated design rules has improved and stabilized over time. This is what I will be turning my attention to next.

References

Ainley, J. (2000). Exploring the transparency of graphs and graphing. In T. Nakahara & M. Koyama, (eds.) *Proceedings of the 24th Conference of the International Group for the Psychology of Mathematics Education*, PME-24 (2-16).

Anderson, L. W. (2005). Objectives, evaluation, and the improvement of education. *Studies in Educational Evaluation*, 31, pp.102-113.

Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Anderson, P. & Morgan, G. (2008). *Developing Tests and Questionnaires for a National Assessment of Educational Achievement.* Washington DC: The World Bank

Ball, S.J. (2003). The teacher's soul and the terrors of performativity, *Journal of Education Policy*, 18:2, pp. 215-228.

Barab, S. A., Evans, M. A., & Baek, E.-O. (2004). Activity theory as a lens for characterizing the participatory unit. In D. H. Jonassen (Ed.), *Handbook of research on educational communities and technology* (pp. 199–214). Mahwah, NJ: Lawrence Erlbaum Associates.

Barab, S. A., Barnett, M., Yamagata-Lynch, L., Squire, K., & Keating, T. (2002). Using activity theory to understand the contradictions characterizing a technology-rich introductory astronomy course. *Mind, Culture, and Activity*, 9(2), pp. 76–107.

Bassey, M. (1999). *Case study research in educational settings*. Buckingham: Open University Press.

Biggs, J. (1998). Assessment and classroom learning: A role for formative assessment? *Assessment in Education: Principles, Policy and Practice, 5(1)*, pp. 103-110.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, n, 5(1)*, pp. 7-74.

Bloom, B. S. (1956). *Taxonomy of educational objectives, the classification of educational goals, handbook I: Cognitive Domain*. New York: David McKay.

Bloom, B.S., Hastings, J.T., & Madaus, G.F. (1971). Handbook on formative and *summative evaluation of student learning*. New York: McGraw Hill.

Boertien, H., & de Lange, J. (1994). *The national option of TIMSS in the Netherlands*.
Enschede, the Netherlands: University of Twente.

Booyse, J.J. (1993). *Research in Education*. Pretoria: University of South Africa.

Braun, H, Kanjee, A, Bettinger, E & Kremer, M. (2006) *Improving Education through Assessment, Innovation, and Evaluation*. American Academy of Arts and Sciences, Cambridge.

Brodie, K., Shalem, Y., Manson, L., & Sapire, I. (2008). *In the interface: mathematics teachers 'finding the curriculum'*. Paper presented at the Kenton Education Association conference, Broederstroom, 26-28 October 2008.

Bryant, S.L., Forte, A., & Bruckman, A. (2005). Becoming Wikipedian: Transformation of a Participation in a Collaborative Online Encyclopedia, *Proceedings of the GROUP international conference on Supporting Group Work*, November, 2005, 1-10.

Buhlungu S, Daniel J, Southall R, Lutchman J, (2007). *State of the Nation. South Africa, 2007*. Cape Town: Human Sciences Research Council.

Chinapah, V. (2003). *Monitoring Learning Achievement (MLA) Project in Africa*, ADEA Biennal Meeting 2003. Grand Baie, Mauritius, 3-6 December.

Chisholm, L (2005). The making of South Africa's national curriculum statement, *Journal of curriculum studies* 37 (2), pp. 193-208.

Clement, J. (1985). Misconceptions in graphing*, Proceedings of the 9th Conference of the International Group for the Psychology of Mathematics Education*, 1, pp. 369 - 375.

Cohen, L. & Manion, L.(1991). *Research Methods in Education*. New York: Routledge.

Cole, M. (1988). Cross-cultural research in the sociohistorical tradition. *Human Development* 3, 1, pp. 137-157.

Crisp, V. & Sweiry, E., 2003. Can a picture ruin a thousand words? Physical aspects of the way exam questions are laid out and the impact of changing them. *Paper presented at the British Educational Research Association Annual Conference, Edinburgh, September 2003*: University of Cambridge Local Examinations Syndicate.

Daniels, H. (2001). *Vygotsky and pedagogy*. London: Routledge Falmer.

Darling-Hammond, L. (2000). No child left behind and high school reform. *Harvard Educational Review*, 76(4). pp. 642-667.

Darling-Hammond, L. (1989). Accountability for professional practice. *Teachers College Record*, 91(1), pp. 55-80.

Davison, C., & Hamp-Lyons, L. (2009). The Hong Kong Certificate of Education: school-based assessment reform in Hong Kong English language education. In L-y Cheng & A. Curtis (Eds.). English Language and the Chinese Learner. Routledge.

de Castro, M.H.G (2000). Education Assessment and Information Systems in Brazil. National Institute for Educational Studies and Research (INEP), Brazil: Ministry of Education.

de Lange, J. (1999). *A framework for classroom assessment in mathematics*. Unpublished manuscript, Freudenthal Institute & National Center for Improving Student Learning and Achievement in Mathematics and Science, Assessment Study Group.

de Mello, V.B. (2011), *Mapping State Proficiency Standards Onto the NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005–2009* (NCES 2011-458). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, DC: Government Printing Office.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). New York: McGraw-Hill.

Department of Basic Education (DBE). (2009). *Report of the Ministerial Task Team for the Review of the Implementation of the National Curriculum Statement*. Pretoria: DBE.

DBE (2010) *Action Plan to 2014: Towards the Realisation of Schooling 2025*. Pretoria: DBE.

DBE (2011a). *Annual National Assessments 2011: A guideline for the interpretation and use of ANA results*. Pretoria: DBE.

DBE (2011b). *Annual National Test Guidelines* [online]. Available at http://www.education.gov.za [accessed 2 September 2013]

DBE (2011c). *Curriculum and Assessment Policy Statement (CAPS) Intermediate Phase Grades 4-6 Mathematics*. Pretoria, DBE

DBE (2011d). *Report on the Annual National Assessments of 2011*. Pretoria: DBE.

DBE. (2012a). *National Protocol for Assessment: Grade R-12*. Pretoria: DBE.

DBE. (2012b). *Report on the Annual National Assessments of 2012*. Pretoria: DBE.

DBE. (2013). *Annual National Assessment Diagnostic Report and 2014 Framework for Improvement.* Pretoria: DBE.

DBE. (2014). *Report on the Annual National Assessment of 2014*: Grade 1 to 6 and 9. Pretoria: DBE.

Department of Education (DoE). (1998). *Assessment Policy in the General Education and Training Band – Grade R to 9 and ABET*. Pretoria: DoE.

DoE (2002a). *Revised National Curriculum Statement* Assessment Guidelines for Mathematics, Pretoria: DoE.

DoE, (2002b). *Revised National Curriculum Statement Grades R-9 (Schools) Policy Mathematics*. Gazette No. 23406, vol. 443. Pretoria: DoE.

DoE (2003a*). Revised National Curriculum Statement Grades R-9 (Schools) Teacher's Guide for the Development of Learning Programmes Mathematics*. Pretoria: DoE.

DoE (2003b). *School enrolment, pass rates up*. Parliament report. http://www.southafrica.info/ess_info/sa_glance/education/education_060603.htm.

DoE (2005a). Grade 6 Intermediate Phase Systemic Evaluation Report, Pretoria: DoE.

DoE (2005b). *National curriculum statement, general education and training assessment guidelines for mathematics (Intermediate and Senior Phase*s). Pretoria: DoE.

DoE (2005c). *National policy on assessment and qualifications for schools in the general education and training band.* Pretoria: DoE.

DoE (2005d). *National protocol on Assessment for Schools in the General and Further Education and Training band (Grades R-12)*. Pretoria: DoE.

DoE (2008a). *Foundations for Learning Assessment Framework Intermediate Phase*, Pretoria: DoE.

DoE (2008b). *Grade 3 Systemic Evaluation 2007 Leaflet*. Pretoria: DoE.

DoE (2009a). Trends in Education Macro Indicators Report. Pretoria: DoE.

DoE (2009b). *Mathematics Examination guidelines Grade 12 2009*, Pretoria: DoE.

Development Bank of South Africa (DBSA) (2008). *Education roadmap, focus on*

schooling system. Pretoria: Development Bank of South Africa.

Emerson, R.W. (1850). *Representative Men, seven lectures. Boston*: Phillips, Sampson and Company.

Engeström, Y. (1987). *Learning by expanding: An activity-theoretical approach to developmental research*. Helsinki, Finland: Orienta-Konsultit Oy.

Engeström, Y. (1993). Developmental studies of work as a test bench of activity theory: The case of primary care medical practice. In J. Lave & S. Chaiklin, (Eds.), *Understanding practice: Perspectives on activity and context (*pp. 64-103). Cambridge: Cambridge University Press.

Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engeström, R. Miettinen, & R.-L. Punamaki (Eds.), *Perspectives on activity theory* (pp. 19-38). New York: Cambridge University Press.

Engeström, Y. (2001). Expansive learning at work: Toward an activity-theoretical conceptualization. *Journal of Education and Work*, 14, pp. 133-156.

Engestrom, Y., & Miettinen, R. (2003). Introduction.. In Y. Engestrom, R. Miettinen, & R. Punamaki (Eds.). *Perspectives on Activity Theory*. United Kingdom, Cambridge University Press.

Even, R. (1998). Factors Involved in Linking Representations of Functions, Journal of

Mathematical Behavior, 17(1), pp.105-121.

Freankel, J and Wallen, N. (1990). *How to design and evaluate research in education*. New York: McGraw-Hill Publishing Company.

Free State Department of Education (FSDoE) (2011). *Provincial Strategy on Learner Attainment. Framework Document.* Bloemfontein: FSDoE.

Fuhrman, S.H. 1999. *The new accountability*. CPRE Policy Briefs, Reporting on issues in education reform. University of Pennsylvania: Consortium for Policy Research in Education (CPRE).

Gauteng Department of Education. 2009. *Johannesburg East district circular 13/2009*. *Foundations for Learning Provincial Improvement Plan 2009–2011*. Johannesburg: Gauteng Department of Education.

Goldstein, H., & Thomas, S.M. (2008). Reflections on the international comparatives surveys debate. *Assessment in Education: Principles, Policy and Practice*, 15(3).215-222.

Greaney, V. & Kellaghan, T. (2008). *Assessing National Achievement Levels in*

Education: The International Bank for Reconstruction and Development; Washington DC: World Bank.

Grønmo, S.L., Lindquist, M., Arora, A., & Mullis, V.S. (2013). *TIMSS 2015 Mathematics*

Framework, Chapter 1, IEA TIMSS and PIRLS International Study Centre: Lynch School of Education, Boston College .

Guba, E. G., & Lincoln, Y. S. (1983). Epistemological and methodological bases of naturalistic inquiry. In G. F. Madaus, M. S. Scriven,& D. L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 311–334). Boston: Kluwer-Nijhoff Publishing.

Hadjidemetriou, C., & Williams, J. (2003). Children's graphical conceptions: assessment of learning for teaching, University of Manchester.

Hart, K. M. (1981). Children's Understanding of Mathematics, 11-16, London: John Murray.

Hess, K.K. (2006). *Exploring Cognitive Demand in Instruction and Assessment*, National Centre for Assessment, Dover. khess@nciea.org.

Hoadley, U. & Jansen, J. (2009). Curriculum, organising knowledge for the classroom, Extract of Section 6 of the work *Curriculum*, SAIDE: Oxford University Press.

Hoadley U (2012) What do we know about teaching and learning in South Afrcan primary schools? Education as Change, 16(2): pp. 187-202.

Howie S & Van Staden S (2012) South African Children's Reading Literacy Achievement– PIRLS and prePIRLS 2011 Summary of the Key results (Media briefi ng). Pretoria: Centre for Evaluation and Assessment

Howie, S, 2006. Assessment for Learning Strategies: Merging theory and practice within SA realities. *Paper presented at the GDE Assessment for Learning Conference*. Johannesburg: University of Pretoria.

Howie, S, 2002. *English Language Proficiency and Contextual Factors Influencing Mathematics Achievement of Secondary School Pupils in South Africa.* Pretoria: PrintPartners Ipskamp.

Husén, T. (1987). Policy impact of IEA research. *Comparative Education Review*, 31, pp. 29-46.

Issroff, K., & Scanlon, E. (2002). Using technology in higher education: An activity theory perspective. *Journal of Computer Assisted Learning*, 18, pp. 77-83.

Jansen, J.D. 2001. “Image-ining teachers: Policy images and teacher identity in South African classrooms*”. South African Journal of education* 21(4): pp. 242-246.

Janvier, C. (1981). Use of Situations in Mathematics Education, Educational Studies in Mathematics, 12, 113-122.

Jaworski, B. and Potari, D. (2009) Bridging the macro- and micro-divide: using an activity theory model to capture sociocultural complexity in mathematics teaching and its development. *Educational Studies in Mathematics* 72, 2, pp. 219-236.

Jurdak, M. (2009). *Toward Equity in Quality in Mathematics Education*, Beirut: Springer.

Kanjee, A. (2006). *The State of National Assessments of Learner Achievement*. Unpublished paper prepared for the Human Sciences Research Council, Pretoria, South Africa.

Kanjee A & Moloi Q. 2014. South African Teachers’ Use of National Assessment Data: *South African Journal of Childhood Education*, *2014 4(2)*:90-113.

Kaptelinin, V.; Nardi, B. A. & Macaulay, C. (1999), 'Methods & tools: The activity checklist: a tool for representing the "space" of context', *interactions* 6 (4), 27-39.

Kerslake, D. (1993). Graphs, in K. Hart (ed.) Children’s understanding of mathematics 11-16, London: John Murray.

Klecker, M, 2008. *Is Teacher Quality Related to Eighth-Grade Mathematics Achievement? Evidence from the 2007 NAEP Data*. Paper presented at the annual meeting of the Mid-South Educational Research Association Knoxville, Tenesee: Morehead State University.

Klinger, D.A., DeLuca, C., & Miller, T. (2008). The evolving culture of large-scale assessments in canadian education, *Canadian Journal of Educational Administration and Policy, Issue #76*, July 3, 2008. pp. 1-31.

Kuiper, W. A. J. M., Bos, K. Tj., & Plomp, Tj. (2000). The TIMSS National Option Test. *Studies in Educational Evaluation*, 26, pp. 43−60.

Kuutti, K. (1996). Activity theory as a potential framework for human-computer interaction research. In B. A. Nardi (Ed.), *Context and consciousness: Activity theory and human-computer interaction* (pp. 17-44). Cambridge, MA: The MIT Press.

Leong, S. C. (2006). *On varying the difficulty of test items*. A paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore, 21-26 May 2006.

Leont'ev, A. N. (1981). The problem of activity in psychology. In J. V. Wertsch (Ed.), The concept of activity in Soviet psychology (pp. 37-71). Armonk, New York: Sharpe.

Leont'ev, A. N. (1974). The problem of activity in psychology. *Soviet Psychology*, 13(2), pp. 4-33.

Lesh, R., & Lamon, S. J. (Eds.). (1992). Assessment of authentic performance in school mathematics (pp. 319–342). Washington, DC: American Association for the Advancement of Science Press.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications.

Linn, K.L., & Miller, M.D. (2005). Planning Classroom Tests and Assessments. *Measurement and Assessment in Teaching*. New Jersey. Pearson Prentice Hall.

Long , C (2006). *Realising the potential of school based assessments*. Paper delivered at the 32nd Annual Conference of the International Association of Educational Assessment, Singapore 21-26 May.

Lunzer, E.A., & Pumfrey, P.D., (1966), "Understanding Proportionality", Mathematics Teaching, 34, 7-12. -Middleton, J., Heuvel-Panhuizen, M. van den, 1995, "The Ratio Table", Mathematics Teaching in the Middle School, 1(4), pp. 282-288.

Maxwell, J.A. (1992). Understanding and Validity in Qualitative Research, *Harvard Educational Review*; Fall 1992; 62, 3; Research Library Core, pp. 279-300.

Merriam, S. B. (1998). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass.

Merriam, S. B. (2002). Introduction to qualitative research. In S. B. Merriam (Ed.), *Qualitative research in practice: examples for discussion and analysis* (1st ed., pp. 1-17). San Francisco: Jossey-Bass.

Meyer, C.B. (2001). A Case in Case Study Methodology, *Field Methods*, Vol. 13, No. 4, November 2001, pp. 329–352.

Misailidou, C. & Williams, J. (2002). *Ratio": raising teachers' awareness of children's Thinking*, Proceedings of the 2nd International Conference on the teaching of Mathematics. http://www.math.uoc.gr/~ictm2/Proceedings/pap143.pdf.

Moloi, M. & Chetty, M. (2011). *The SACMEQ III Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education*. Pretoria: DBE.

Moloi, M. & Strauss, J. (2005). *The SACMEQ II Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education*. Harare: SACMEQ.

Morgan, C. (2000). *Discourses of assessment – Discourses of mathematics*. Paper presented at the conference of Mathematics Education and Society 2. Montechoro, Portugal.

Nardi, B. (2007). Placeless organizations: Collaborating for transformation. *Mind, Culture and Activity: An International Journal*, 14(1/2). pp. 5-22.

New England Common Assessment Program (NECAP) (2013). *Vermont's Framework of Standards*. Vermont: Vermont Agency of Education.

Nunez, I. (2009). Activity Theory and the Utilisation of the Activity System according to the Mathematics Educational Community, *Educate~ Special Issue*, *December 2009*, pp 7-20. http://www.educatejournal.org/.

Ogomaka, P.M.C., Nosike, M.C., & Akukwe, A.C. (2013). Task Number and Cognitive Complexity as Determinants of Difficulty Levels In Multiple – Choice Test Items, *West African Journal of Industrial and Academic Research* Vol.8 No.1 September 2013, pp. 227-235.

Osborne, R. and Gilbert, J. (1980). A technique for exploring student's views of the world. *Physics Education (15)*, pp. 376-379.

Pellegrino JW, Chudowsky N & Glaser R. 2001. *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.

Pereira-Mendoza, L. & Mellor, J. (1990): Students' Concepts of Bar Graphs-Some Preliminary Findings, *ICOTS* 3, pp. 150-157.

Porter, A.C. (2002). Measuring the content of instruction: Uses in research and practice. *Educational Researcher*, 31(7), 3-14.

Posner, G. J. & Gertzog, W. A. (1982). The clinical interview and the measurement of conceptual change*. Sci. Ed*., 66: pp. 195–209.

Post, T., Cramer, K., Behr, M., Lesh, R., & Harel, G. (1993). Curriculum implications of Research on the Learning, Teaching, and Assessing of Rational Number Concepts. In T. Carpenter & E. Fennema (Eds.), *Research on the Learning, Teaching, and Assessing of Rational Number Concepts*: Lawrence Erlbaum and Associates.

Postlethwaite, T.N., and Kellaghan, T. (2008): *National assessments of educational achievement, Education Policy Series* (9), UNESCO, Paris: International Institute for Educational Planning

Ramphele, M. (2009). *Another generation betrayed*. The Sunday Times (January 18, 2009), p.19.

Rao, D.B & Harshitha, D. (2004). *Assessing Learning Achievement*. New Delhi: Discovery Publishing House.

Reddy V, Prinsloo C, Visser M, Arends F, Winnaar L, Rogers S et al. (2012). *Highlights from TIMSS 2011: The South African perspective*. Pretoria: HSRC.

Reddy, V., 2006. *Mathematics and Science Achievement at South African Schools in TIMSS 2003*. HSRC Press, Cape Town.

Reeves, C. & Muller, J., 2005. Picking up the pace: variation in the structure and organization of learning school mathematics. *Journal of Education. 37 (1)*: pp. 1–28.

Robitaille, D. F., Schmidt, W. H., Raizen, S., Mc Knight, C., Britton, E., & Nicol, C. (1993). *TIMSS Monograph No. 1. Curriculum framework for mathematics and science*. Vancouver,BC, Canada: Pacific Educational Press.

Ross, K. N & Genevois, I. J. (2006). Cross-National Studies of the Quality of Education: Planning Their Design and Managing Their Impact, Paris: International Institute for Educational Planning.

Ross, K. (2005). *Quantitative Research Methods in educational planning*. Paris International Institute for Educational Planning/UNESCO.

Roth, W.M., & Tobin, T. (2002). Redesigning an “urban” teacher education program: an activity theory perspective*. Mind, Culture and Activity*, 9(2), pp. 108-131.

Sannino, A., Daniels, H., & Gutiérrez, K.D. (2009). *Learning and Expanding with Activity Theory*, New York: Cambridge University Press.

Scaife, J. (2004). Reliability, Validity and Credibility. In C. Opie (Ed.), *Doing Educational Research: A Guide to First Time Researcher* (pp. 58-72). London: Sage.

Schumacher, S & McMillan, J. (1993). *Research in Education*. New York: Harper Collins College Publishers.

Scriven, M. 1983. "Evaluation ideologies". In Madaus, G. F.; Scriven, M.; and Stufflebeam, D. L. (eds.). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Norwell, Massachusetts: Kluwer. pp. 229-260.

Sharma, S. S. (1993). *Graphs: Children's strategies and errors*. London: Kings College.

Sisouk, V., & Postlethwaite, T. N. (2007). Laos grade 5 national assessment. Vientiane, Laos: Ministry of Education.

Smith J, Dietiker, L, Lee, K, Males, LM, Figueras, HM, Mosier, A, Chang, KL*, & Sisman, G, (2008). Paper prepared for the 2008 Annual Meeting of the American Educational Research Association, New York: STEM Project.

Soudien, C. (2008). *The implications of the crisis in numeracy and literacy in South Africa for teacher education*. Centre for Education Policy Development Annual Report 2008. Johannesburg: CEPD.

Spaull, N. (2013). *South Africa's education crisis*. Johannesburg: Centre for Development and Enterprise (CDE).

Spaull, N. (2015). Accountability and capacity in South African education, *Education as Change, 2015 19(3)*, 113–142.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks: Sage.

Steen, L.A. (1990). On the Shoulders of Giants: New Approaches to Numeracy. Washington, DC: National Academy Press: Stein and Smith (1998 / 2009).

Stein, M.K., Smith, M.S., Henningson, M.A. and Silver, E.A. (2000). *Implementing standards-based mathematics Instruction, A casebook for professional development*, New York: Teacher College Press.

Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83(10)*, pp. 758-765.

Swaffield, S & Thomas, S (2016) Educational Assessment in Latin America, Assessment in Education: *Principles, Policy & Practice*, 23:1, pp.1-7.

Taylor, N. (2008). *What's wrong with South African schools?* Retrieved from JET Education Services: www.jet.org.za.

Taylor, N. and Vinjevold, P. (1999). *Getting Learning Right: Report of the President's Education Initiative Research Project*. South Africa: Joint Education Trust.

Thompson, P. (1978). *The voice of the past*. Oxford: Oxford University Press.

UNESCO. (2005). *Education For All Global Monitoring Report 2005*. Paris: UNESCO Publishing.

Van Daler, D. B. (1973). *Understanding Educational Research – An Introduction*. McGraw Hill Inc.

van der Berg, S. (2005). *How effective are poor schools? Poverty and educational*

outcomes in South Africa. Paper presented to the SACMEQ conference, Paris, September.

van der Berg, S., & Louw, M. (2006). *Unravelling the mystery: Understanding South African schooling outcomes in regional context*. Paper presented at the conference of the Centre for the Study of African Economics, Oxford University.

van der Berg S, Burger C, Burger R, de Vos M, du Rand G, Gustafsson M, Moses E, Shepherd D, Spaull N, Taylor S, van Broekhuizen H & von Fintel D (2011) *Low Quality Education as a Poverty Trap*. Stellenbosch: University of Stellenbosch.

Van Oers, B. (2001). Educational forms of initiation in mathematical culture. *Educational Studies in Mathematics*, 46, pp. 59-85.

Vandeyar, S., & Killen, R. 2003. Has Curriculum reform in South Africa really changed assessment practices, and what promise does the revised national curriculum statement hold? *Perspectives in Education*, 21(1). pp. 119-134.

Venkatakrishnan, H. (2004). *The implementation of the mathematics strand of the key stage 3 strategy: A comparative case study*. Unpublished PhD thesis, London, King's College.

Venkat, H, & Adler, J. (2012). Coherence and connections in teachers' mathematical discourses in instruction*. Pythagoras*, 33(3), 25–32.

Verenikina,I. & Gould, E. (1998) Cultural-historical Psychology & Activity Theory. In Hasan, H., Gould., E. & Hyland, P. (Eds.) *Activity Theory and Information Systems* (7-18), Vol. 1.Wollongong: UOW Pres.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Watson, A. (2000). Mathematics Teachers Acting as Informal Assessors: Practices, Problems and Recommendations. *Educational Studies in Mathematics (41)*, pp. 69-91.

Webb, N.L. (2002): Depth-of-Knowledge Levels for Four Content Areas, unpublished paper. http://facstaff.wcer.wisc.edu/normw/All.doc.

Wertsch, J. V. (1991). *Voices of the mind: A Sociocultural approach to mediated action*. Cambridge, Massachusetts: Harvard University Press.

Westberry, N.C. (2009). *An activity theory analysis of social epistemologies within tertiary-level eLearning environments*. Doctoral Thesis, University of Waikato:

Research Commons. http://waikato.researchgateway.ac.nz/.

Williamson, J. B. (1982). *The research craft*. Canada: Little, Brown & Company Ltd.

Withers, G. (2005). *Item Writing for Tests and Examinations: Module 5*. Paris: International Institute for Educational Planning/UNESCO.

Annexures

Appendix 1: Principal consent letter and form

Enquiries: M Chetty [redacted] Chetty.m@doe.gov.za

M Chetty
[redacted]

The Principal
[redacted]

Dear Sir

I am currently a part time student studying for a doctoral degree in education through the University of the Witwatersrand. My area of interest involves exploring interacting assessment systems in Mathematics: comparing school-based assessments to standardised national assessments. My supervisor at the university is Prof ….

I have obtained formal approval from the Gauteng Department of Education (GDE) to conduct research in five primary schools. I have selected your school as one of the five schools to conduct my study. The targeted research participants are all the grade 6 mathematics teachers at your school. The data collection will focus on teachers' assessment tools and planning, a sample of learner scripts/test books and any relevant documents/guidelines teachers use to design their internal formal assessment tasks. With all documentation retrieved, the anonymity of the school and the teachers will be maintained and will only be considered for the purposes of this research study.

It is with great enthusiasm that I look forward to conducting the research at your school and involving the grade 6 mathematics teachers as willing research participants. You can be assured that my research activities at the school will not interfere with teachers' normal workloads and will in general limit my interaction with them to non-teaching hours at the school. Attached to this letter is a consent form regarding my data collection plan with intended activities and timeframes at the school. I hope you will find this in order.

Kind regards

Mark Chetty

Principal consent from

Overview of the data collection plan

ACTIVITY	TOOL	TIMEFRAME	PARTICIPANTS	PURPOSE
Conduct an informal introductory meeting at each school	Journal	Feb	Principal and grade 6 mathematics teachers at each school	To discuss purpose of the study and submit research request letters to the principal and teacher participants.
Profile school and selected teachers	School profile form and teacher profile form	Feb	Principal and grade 6 mathematics teachers at each school	To establish individual teacher profiles and socio-economic status of school.
Collect grade 6 mathematics assessment documents.	Checklist for collecting documents	Feb	Grade 6 mathematics teachers at each school	To collect information for document analysis
First interview session with teachers	Interview schedule	May		Conduct individual interviews with teachers
Second interview session with teachers	Interview schedule and questionnaire	August		Conduct reflective interviews with teachers
Collect questionnaires from teachers and meet with principal	Journal	August		To establish feedback from teachers on data collection process

Research request is approved

Name of school:	
Principal:	
Signature:	Date:
School Stamp:	

Appendix 2: Teacher consent letter and form

Enquiries: M Chetty [redacted] Chetty.m@doe.gov.za

M Chetty
[redacted]

The Teacher
School:
Setting (SES):
1455

Dear Mr/Mrs/Miss _____

Following consent from the principal regarding the participation of grade 6 mathematics teachers at the school to conduct my research, I am submitting this letter to formalize the process. I am currently a part time student studying for a doctoral degree in education through the University of the Witwatersrand. My area of interest involves exploring interacting assessment systems in Mathematics: comparing school-based assessments to standardised national assessments. My supervisor at the university is Prof ….

The data collection will focus on your assessment tools and planning, a sample of learner scripts/test books and any relevant documents/guidelines used to design your internal formal assessment tasks. With all documentation retrieved, the anonymity of the school and yourself will be maintained and will only be considered for the purposes of this research study.

It is with great enthusiasm that I look forward to conducting the research at your school and involving you as a willing research participant. You can be assured that my research activities at the school will not interfere with your normal workloads and will in general limit our interactions to non-teaching hours at the school. Attached to this letter is a consent form regarding my data collection plan with intended activities and timeframes at the school. I hope you will find this in order.

Kind regards

Mark Chetty

Teacher consent from

Overview of the data collection plan

ACTIVITY	TOOL	TIMEFRAME	PARTICIPANTS	PURPOSE
Conduct an informal introductory meeting at each school	Journal	Feb	Principal and grade 6 mathematics teachers at each school	To discuss purpose of the study and submit research request letters to the principal and teacher participants.
Profile school and selected teachers	School profile form and teacher profile form	Feb	Principal and grade 6 mathematics teachers at each school	To establish individual teacher profiles and socio-economic status of school.
Collect grade 6 mathematics assessment documents.	Checklist for collecting documents	Feb	Grade 6 mathematics teachers at each school	To collect information for document analysis
First interview session with teachers	Interview schedule	May		Conduct individual interviews with teachers
Second interview session with teachers	Interview schedule and questionnaire	August		Conduct reflective interviews with teachers
Collect questionnaires from teachers and meet with principal	Journal	August		To establish feedback from teachers on data collection process

Research request is approved

Name of school:	
Teacher:	
Signature:	Date:
School Stamp:	

Appendix 3: First Interview schedule

Interview with Kalay – Suburban School
Date:

RQ 1: justifications for designing and selecting assessment tasks

1. Introduction

Good morning (Teacher X), thank you for making time for this interview. As I discussed in our telephone conversation the other day (26/11/2010) that I would be conducting an interview with you on issues emanating from the document analysis (your assessment files). This interview is mainly about understanding more about the aspects that figure in your design (form) of assessments. Please feel 'free' to discuss matters openly as I'm here as a researcher interested in assessment issues not as an official looking at policy compliance.

Questions

a) General

1. Can you describe briefly your teaching background, subjects and grades you teach, involvement in school structures (e.g. those dealing with assessment), etc.

b) National Policy texts

1. Some of the policy documents you have in your file include the NCS and there are documents that refer to the milestones assessment framework of the FFL. Can you tell me how you use these documents for assessment planning?

P1: Have you been trained/work-shopped on using these documents to plan for assessments.

P2: Are these documents easy to use?

P3: How is the milestone curriculum used in your planning?

P4: Do you rely/use any textbooks to assist you make sense of the national documents?

P5: How do you feel about the current packaging (formulation) of the current curriculum in terms of it assisting you with assessments?

c) District level mediation

2. In addition to the national policies/frameworks, you have a district documentation relating to assessments in your file (invites, notices, cluster meetings, workshops). Can you tell me more about some of the district activities you attend or are involved with?

P1: How often do you attend district meetings?

P2: Are some of these meetings held specifically for grade 6 teachers?

P3: To what extent does cluster moderation take place (ref: twinning of schools)?

P4: What takes place at these at these cluster moderation meetings?

P5: How do you feel about the feedback you receive at these meetings?

d) School Assessment

Following on the various school assessment documents you have in your file, there is evidence of different types of planning (learning programme, work schedule and lesson plans) in your files.

3. Can you tell me more about how you structure you're planning.

P1: How is assessment planning featured at each level? Do you plan individually or as a collective within the mathematics department (ref: composite assessment plan).

P2: Can you explain these different types of planning using one topic (e.g. ratio)?

P3: The compilation of an assessment task in your recording sheet is quite involved with various sub-activities. Can you share more information on this?

P4: How do you go about deciding on these sub-activities?

P5: Is there any assessment type that you prefer to administer to your learners?

P6: How do you go about deciding on these sub-activities?

P7: Do you format your own tests (typing, printing …)

P8: How do you handle the selection of contexts in your assessments? Do you prefer low language contexts?

P9: Can you explain the difference between control tests and class tests … and mental tests? How are these different from speed tests?

P10: How do you go about giving assignments and investigations?

P11: How is homework assessed?

P12: Your test results across classes indicate very good learner performance (ratings 3 or 4). Are there any particular reasons for this?

e) Assessment design

1. How do you cater for learners with different ability levels in your assessments (ref: symbol analysis)?

2. What do you consider important when designing your assessments?

3. Has your assessments forms/content coverage changed between 2009 and 2010? Can you tell me a little about this and show me some examples on why you introduced these changes. What is your view on these changes?

4. What is opinion about the annual national assessments (ANA) as a standardised tool for assessment?

f) Aspects for next interview:

- copies to be collected on the following topics (2010 assessments).

Table: Coverage of learning outcomes

LO	Topic
1 – Numbers and operations	Problem solving involving appropriate contexts (e.g. financial maths, Rate and ratio).
2 – Number patterns and relationships	Problems involving geometric and numeric number patterns (non-routine problem solving).
3 – Shape and space	Recognition of 2-D shapes and 3-D objects
4 - Measurement	Applications involving time
5 – Data handling	Problems on data handling from bar graphs and pie charts

g) Conclusion

Thank you (Teacher X) for your time and sharing of information. I will be in touch with you regarding our next interview which will focus mainly on the mathematics content you include in your assessments.

Appendix 4: Coding of formal examination task

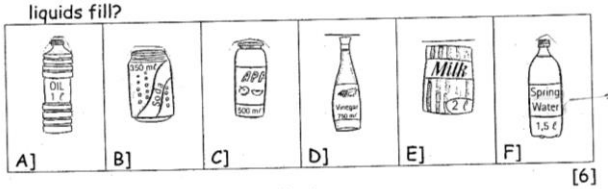
a) Interview questions:

1. How is the test developed/designed and for what purpose?
2. How do you feel about the standard of the test?
3. What categories (or framework) do you use in the design?
4. How do you decide on the assessment standard to assess?
5. How did you decide on the number of questions and sub-questions?
6. Is there a build-up of sub-questions?
7. What do consider being the intended difficulty of a question?
8. How do you deal with the language and real world contexts used?
9. How do you incorporate different thinking skills (e.g. knowledge type/ routine) in your tasks?
10. How do you allocate marks?
11. What do you consider the main influence in designing the test?

b) Coding table on the range and scope of questions

The assessment task is first considered in light of the interview data held with the teacher. Thereafter, the teacher is analysed using the specific descriptors on range and scope, drawn from the literature in the study. The exercise is repeated on sample extracts from formal examinations, class tests and smaller activity based assessments to establish a specific story of the teachers' SBA tasks.

The table below shows a coded extract from a formal examination task of Kalay from the suburban school.

<table>
<tr><td>Teacher</td><td>Sample extract from formal examination</td><td colspan="5">Interview context</td></tr>
<tr><td>Kalay</td><td><u>Question Three: Measurement [Length]</u>
3.1 Complete these sentences. The words are in the list.
thousand hundred ten million
1. One km is one _____ m.
2. There are _____ mm in one cm.
3. One m is one _____ mm.
4. There are one _____mm in one km. [4]
3.2 If a glass holds 250ml of liquid, how many glasses can these liquids fill?

A] B] C] D] E] F] [6]
3.3 Problem Solving → Word Problems
At athletics meeting in the shot-put event, each competitor is allowed three throws. These boys scored as follows:
<table><tr><th></th><th>Pontsho</th><th>Aadil</th><th>Siyabonga</th></tr><tr><td>1st throw</td><td>6,456 m</td><td>6,972 m</td><td>6,439 m</td></tr><tr><td>2nd throw</td><td>7,009 m</td><td>6,673 m</td><td>7,126 m</td></tr><tr><td>3rd throw</td><td>6,875 m</td><td>7,101 m</td><td>7,102 m</td></tr></table>a] Who was the winner?
_____ [2]</td><td colspan="5">Kalay indicated that there was always an attempt to design mathematics tasks with varying ability levels. Her selection of questions was primarily based on work covered in class work and would be familiar to her learners. She prefers to set her own assessments than rely on the externally set papers. There is however more than one level of moderation that her tests are subjected to and she values the inputs made within her school. She did acknowledge that most of her questions was not her original work but were carefully selected from suitable texts. She acknowledged that she was not always certain about the difficulty or cognitive demand of a question and seemed to have some implicit notion of differentiation. Although there is a strong within school community that looks at the pragmatics of whether most learners will cope with the questions, there is not a focused support on the design of an appropriate test using a pre-determined framework.</td></tr>
<tr><td>DL</td><td>Are the knowledge elements within the scope of the grade?</td><td>E – below grade</td><td>M – within grade</td><td>D – above grade</td><td></td><td>Knowledge elements are within scope of grade – mainly moderate difficulty with some top-end working required in 3.3.</td></tr>
<tr><td>CD</td><td>Is there a range of cognitive demand of questions?</td><td>K</td><td>R</td><td>C</td><td>N</td><td>Range builds from knowledge and routine into complex problem solving</td></tr>
<tr><td>LD</td><td>What is the language demand?</td><td>NT</td><td>LT</td><td>HT</td><td></td><td>Range has low text; but mainly high text questions in the</td></tr>
<tr><td>IF</td><td>What is the format of the items?</td><td colspan="2">CSA</td><td>MCQ</td><td></td><td>All items are closed short answer type with varying formats utilising words in blocks, pictures in blocks and data in tables.</td></tr>
</table>

Appendix 5: List of Acronyms

ANA	Annual National Assessment
AS	Assessment Standard
AT	Activity Theory
AT	Activity Theory
CAPS	Curriculum and Assessment Policy Statements
CD	Cognitive Demand
CSA	Closed Short Answer
DBE	Department of Basic Education
DL	Difficulty Level
DoE	Department of Education
DOK	Depth of Knowledge
EMT	Expanded Mediation Triangle
FFL	Foundations For Learning
GDE	Gauteng Department of Education
GET	General Education and Training Band
HT	High Text
IEA	International Association for the Evaluation of Educational Achievement
IF	Item Format
IIEP	International Institute for Educational Planning
IRT	Item Response Theory
LD	Language Demand
LO	Learning Outcome
LT	Low Text
LTA	Learning, Teaching and Assessing
MCQ	Multiple choice question
MD	Mathematical Domain
MLA	Monitoring Learning Achievement
NAEP	National Assessment of Educational Progress
NCS	National Curriculum Statement
NCTM	National Council of Teachers of Mathematics
OBA	Outcomes Based Assessment
OBE	Outcomes Based Education
OECD	Organization for Economic Co-operation and Development
PED	Provincial Education Department
PIRLS	Progress in International Literacy Study
RNCS	Revised National Curriculum Statement
SACMEQ	Southern and East Africa Consortium for Monitoring Educational Quality
SBA	School Based Assessment
SNA	Standardised National Assessment
SBST	School-based support team
SE	Systemic Evaluation
SES	Socio-Economic Status
SNA	Standardised National Assessment
TIMSS	Trends in International Mathematics and Science Study

Appendix 6: Definition of terms

Term	Meaning
Annual National Assessment	This refers to the administration of low stakes nationally standardised tests in Mathematics and Languages, targeting grades to be determined by the Minister of Basic Education in South Africa.
Assessment	Assessment is the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving learning and development. (Palomba and Banta, 1999). It may also be defined as a process of collecting, analysing and interpreting information to assist teachers, parents and other stakeholders in making decisions about the progress of learners (DBE, 2012). It is a process of reasoning from evidence.
Assessment activities	Assessment activities are activities given to learners to find out what they know and can do. An assessment activity is one in which the teacher is checking to see if learners have met the objectives of the syllabus, lesson or curriculum. Children often learn a lot from good assessment activities. Examples of assessment activities are writing a story or paragraph, making a model, solving problems and role playing.
Assessment standards	Assessment standards reflect the skills, knowledge and values (SKVs) required to achieve learning outcomes. Assessment standards describe the level at which learners should demonstrate their achievement of the learning outcome(s) and the ways (depth and breadth) of demonstrating their achievement. They are grade specific and show how conceptual progression will occur in the learning Area/subject.
Examination	Examination is a formal assessment conducted at the end-of-term and/or a once-off end-of-year assessment (DBE, 2012). Apart from knowing what grade they got, students do not often get feedback on their performance on the examination. Examinations are usually written in the same way that tests are written. They often have important consequences for students' future.
External assessment	This refers to any assessment activity, instrument or programme where the design, development and implementation has been initiated, directed and, coordinated by Provincial Education Departments and the National Ministry of Education either collectively or individually.
Formal Assessment	This means a systematic way of assessment used by teachers to determine how well learners are progressing in a grade and in a particular subject (DBE, 2012). Examples of formal assessments include tests and examinations.
National Assessment	A national assessment may be defined as an exercise designed to describe the level of achievements, not of individual students, but of a whole education system, or a clearly defined part of it (Kellaghan, 2004). It can also be seen as the gathering of relevant information from an education system to monitor and evaluate the performance of learners and other significant role-players as well as the functioning of relevant structures and programs within the system for the purpose of improving learning (Kanjee, 2007).

Question	A test item is usually structured in the form of a question. Good questions are clear and unambiguous, seeking a specific response from a learner.
SBA	This means all formal assessment, including examinations, conducted by the school throughout the year on a continuous basis. These assessments are usually devised and administered by class teachers, although some are the work of the school principal or other teaching staff. Generally, SBA is aligned with the delivered curriculum and may employ a broader array of media (e.g. oral presentations) and address a greater range of topics than is the case with SNA (Braun and Kanjee, 2006). SBA is collected using various forms of assessment, including tests and examinations.
SBA activity system	An activity system designed to show how advocated rules within the school environment give shape to the kinds of assessment tools (artefacts) found within SBA. The system is driven by contradictory relationships between goals, rules and tools leading to disjunctures and tensions in the tools teachers use to assess learners at school level.
SNA	This refers to national or cross-national assessments that aim to provide systemic data about the achievement of learning outcomes defined by the State or agency (either local or international) implementing them. It describes the level of achievements, not of individual learners, but of a whole education system, or a clearly defined part of it (Kellaghan, 2004). It is often used to monitor and evaluate the performance of learners and other significant role-players as well as the functioning of relevant structures and programs within the system for the purpose of improving learning (Kanjee, 2007). TIMSS, SACMEQ, PIRLS and the ANA are examples of SNA studies.
SNA activity system	This refers to a system that involves analysing underlying motives and goals of State agencies within advocated rules meant to influence the design of national assessment tools. The system is characterised by multiple goals linked to varied assessment discourses leading to contradictions and tensions between advocated rules and enacted tools.
Teacher portfolio	This means the recording and planning documents used by the teacher, namely the formal programme of assessment, evidence of learner assessment/performance, all formal assessment tasks and marking guidelines, annual teaching plan/work schedule, textbook used and other resources.
Test item	A test item is used to draw a response to a series of questions or prompts that can be used as evidence of a learner's level of knowledge, competence, or understanding. Good items are clear, relevant to the curriculum, and focused on one aspect of learning. They provide engaging, genuine tasks that are fair to learners of different language and cultural backgrounds. (Anderson and Morgan, 2008).

Tests	The term “test,” as we have defined it, denotes an instrument of assessment that is conducted under some set of formal conditions (Braun and Kanjee, 2006). Tests usually come at the end of a topic or unit to find out what a student has learned. Tests can include a wide range of question types, but the most common are multiple choice, true and false, essays and matching. Learner test scores provide clear information to learners (and their parents) which areas of work they have mastered and which not (Sloane & Kelley, 2003).