Developing a theoretical rationale for the attainment of greater equivalence of standard in the Grade 12 Home Language exit-level examinations

Colleen Lynne du Plessis

Thesis submitted in fulfilment of the requirements for the degree *Doctor Philosophiae* in the Faculty of the Humanities (Department of Linguistics and Language Practice), University of the Free State

Promotor: Prof. A.J. Weideman

February 2017

Table of Contents

List of tables

List of figures

Acknowledgements

This thesis is the culmination of an initial enquiry into the standard of the Grade 12 Home Language (HL) examination that was commissioned by the Council for Quality Assurance in General and Further Education and Training (Umalusi) in 2012 and further expanded into a full research study in the form of a doctoral thesis. I wish to extend my gratitude to Umalusi and Mr Biki Lepota in particular for bringing to my attention the intricacies of the school-leaving language examination and for granting me the opportunity to participate in the compilation of a report on the HL examination papers in collaboration with other members of the Inter-institutional Centre for Language Development and Assessment (ICELDA).

My heartfelt gratitude is further due to the many family members, friends and colleagues who encouraged me to pursue my studies. The role of Prof. Albert Weideman merits special mention. It was through his encouragement and expertise as a renowned applied linguist and language testing professional that I first developed an interest in language assessment. I have had the privilege of being mentored by a person of Prof. Weideman's calibre and have benefited immensely from his vast repertoire of knowledge. I also expressly wish to thank Prof. Bernard Spolsky for his insightful comments on the focus of my research.

Through studying the Grade 12 language examination papers I have come to appreciate the significance of fundamental design principles in applied linguistic artefacts that are used to decide the fate of students who are required to complete distinct types of language assessments in order to graduate from school or college. The enigmatic nature of language learning and the near impossibility of setting and justifying levels of proficiency in a fair and credible manner when assessing language ability continue to intrigue me.

“Act justly, show mercy, walk humbly”

Declaration

I herewith declare that this thesis, which is being submitted to meet the requirements for the degree *Doctor Philosophiae* in the Faculty of the Humanities (Department of Linguistics and Language Practice) of the University of the Free State, is my own independent work and that I have not previously submitted the same work for a qualification at another institution. I agree to cede all rights of copy to the University of the Free State.

1 Assessing Home Language (HL) ability in the Grade 12 external examination

1.1 Introduction

South Africa is a democracy still fraught with inequality. The country's disparate basic education sector provides one of the most patent examples of how unequal treatment can entrench forms of social injustice and impede development. Of the enormous annual intake of well over a million Grade 1 entrants, less than half manage to remain in the school system and progress to Grade 12 level, and of those who do, only around 36% pass their matriculation year (Solidarity Research Institute 2015). Of the fortunate students who go on to matriculate, only a select few would have had the benefit of a satisfactory school education that would place them in a favourable position to pursue a career or profession of their choice. The remainder are kept trapped in a tragic and unfair cycle of semi-literacy/illiteracy, poverty and deprivation. The situation has reached a critical point in the history of the democratic nation and access to equitable and quality education in South Africa has become a burning issue.

There are regular calls for transformation of basic and higher education and campaigns for equal opportunity to study further (and that without having to make a financial contribution) are all too often accompanied by violent and disruptive protests, which further obstruct learning. This disconcerting state of affairs on the educational front frames the subject of the thesis. With its focus on the endeavour to attain equivalence in respect of the way school language subjects are assessed and treated, the study has as its objective working towards a feasible and more comparable language teaching dispensation that will create fair and meaningful learning opportunities for more students. The democratic and constitutional prerogatives of mutual respect and equality of person can only be realised through fair and equitable treatment, i.e. equivalence in as many forms and on as many levels as possible, including the Grade 12 school-leaving examination.

It would be wrong to assert that government has done nothing to improve education. Credit is due where tangible efforts have been made to eliminate discrepancies in terms of infrastructure, funding and educational standards. However, despite several changes to

the curriculum and ongoing attempts to redress inequalities in the school system, the standard of education in South Africa continues to elicit much criticism. Each year the country spends in excess of R200 billion per annum on education, which is approximately 20% of its entire budget, and this amount is projected to increase annually by around 6% (*BuaNews* 2012; Spaull 2012). Yet, we see that many learners continue to leave the system without completing their basic years of schooling, those who do manage to matriculate battle to find employment, and others who proceed to study at higher institutions of learning struggle to pass (Chisholm 2005; John 2012; Parker 2012; Solidarity Research Institute 2012, 2015). This means that a considerable number of learners are neither acquiring the knowledge nor developing the abilities that they need to succeed both during their school years and after leaving school, a matter that warrants urgent investigation.

In view of the strong mediating role that language plays in the teaching and learning process, it is hypothesised that the school language curriculum and assessment protocol could be harnessed to a far greater extent to help students to do better. It is imperative that every effort is taken to ensure that learners benefit from the language curriculum and that their language proficiency and knowledge are of a sufficiently high level to facilitate learning. Without strong language skills, progress in other fields of study will be undermined, as students will not be able to engage properly with learning content. Cognisance thus needs to be taken of the heuristic role that language fulfils in either assisting or obstructing learners in their attempts to access information, negotiate meaning, gain understanding and communicate any newfound knowledge (Uccelli & Snow 2010; Du Plessis 2016). Apart from the importance of a comprehensive language curriculum, assessment is necessary to verify whether adequate learning has taken place. The quality of the instruments used to measure language learning – in the case of the present study, the language examination papers – is thus just as important as the quality of the language education.

Logically speaking, if language is considered to be instrumental in facilitating learning in all subject fields, considerably more attention should be devoted to the language component of the curriculum and to what extent learners/students are becoming proficient in the respective language subjects, so as to increase their chances of knowledge acquisition in other fields of study too. It is therefore disturbing to note how little attention

has been devoted by the education authorities to the Home Language (HL) component of the curriculum. With the exception of English as First Additional Language (FAL), very little research has been commissioned by the education authorities on the standard of the language subjects and the system of assessment. In response to this unsatisfactory situation, this study is aimed at initiating a series of immediate steps to prioritise and valorise the HL subjects by exploring how the socially responsible examination of language ability can potentially be used to articulate the appropriate standards and create as well as sustain a more equitable education system. Though the issues are big, and concern the heart of a democratic dispensation, the steps that need to be taken to remedy injustice and secure constitutionally entrenched promises are, by comparison, often small and incremental. Nonetheless, I hope that this study will begin to make a contribution to justice and equality, by proposing a set of assessment practices that would make life more fair and equal.

1.2 Preamble to the study

The performance of matriculants in the exit-level examination is used to determine which learners will be granted access to tertiary education and who may qualify for financial grants and bursaries to study further. The school-leaving examination thus serves as an important barometer of learning in the classroom on the basis of which inferences are made about the potential progress and ability of learners. Whether the current examination papers serve as reliable and credible indicators of knowledge acquisition and ability, however, is the subject of much ongoing debate.

More than a decade of research into the school curriculum has been commissioned by the Council for Quality Assurance in General and Further Education and Training (Umalusi), the statutorily mandated overseer of the matriculation qualification, and numerous revisions to curriculum statements and assessment practices have been introduced by the Department of Education (subsequently to 2009 referred to as the Department of Basic Education) since 1994. Notwithstanding all these efforts to ensure an equitable and quality education for all South African learners, confidence in the matric examination system continues to be eroded by the perceptions of the public and other sectors that standards are simply far too low (Van Wyk 2012; *The Economist* 2012; Modisaotsile 2012; Solidarity Research Institute 2015). Further to this, accusations of discrepancies

between the standards of the respective examination papers and of inflating the results have been levelled against the education authorities (Wilkinson 2014; John 2012). The high pass rates for some of the HL and FAL subjects (Department of Basic Education 2012d), generally above 94%, are particularly worrisome and only serve to fuel the fire and arouse further suspicion.

An overview of research commissioned by Umalusi shows that the matter of whether the existing requirements for the school-leaving Senior Certificate (SC) fulfilled the requirements of higher education was raised in 2004 already (Umalusi 2011b: 27). Admittedly, most learners will pursue other options after school and only a few will progress to tertiary level. Still, the objective should be to enable learners to advance as far as possible in all areas of learning, not only to achieve a basic functional literacy. The newly introduced Curriculum and Assessment Policy Statement (CAPS) for the language subjects certainly makes it clear that one of the objectives of the HL curriculum is to prepare learners for tertiary study (Department of Basic Education 2011a).

The introduction of the National Benchmark Tests (NBTs) as part of the admissions process at many South African universities (CETAP 2012) has also contributed towards the lack of confidence placed in the results of the National Senior Certificate (NSC, previously SC). Higher Education South Africa (HESA) commissioned the development of these tests in 2005 to assess proficiency in Academic Literacy, Quantitative Literacy and Mathematics, and as a means of interpreting the results of school-leaving examinations such as the NSC. By compelling university applicants to write the NBT, the impression is created that the results of the NSC cannot be trusted. This provides a strong impetus for undertaking a comprehensive study to validate the system of assessment and HL examination papers, and determine whether there are grounds for scepticism.

Any study of the assessment of language ability in an examination context will have to reflect and relate this assessment to current paradigms in language teaching and testing. Several decades have passed since the advent of communicative language teaching and recognition of the need to relate the measurement of performance to the use of language in authentic social settings (Bachman & Palmer 1996; Plakans 2012; Shin 2012; Young 2012). The structuralist and restrictive view of language as a combination of sound, form and meaning (phonology, morphology, syntax and semantics), necessitating the

assessment of separate skills, is no longer advocated. If it is then argued that language is always used in an integrated manner in specific contexts, and that skills simply cannot be isolated momentarily for assessment purposes, it is surprising to note the continued categorisation of assessment artefacts as “reading”, “listening”, “speaking” or “writing” examinations. Despite the recognition of the failure of previous paradigms of language assessment to take into account the communicative role of language as a social instrument used to mediate and negotiate interaction in a variety of specific contexts (Van Dyk & Weideman 2004a; Riley 2010; Young 2012), there is still little evidence of a move towards the design of integrated language tasks (Stoynoff 2009; Plakans 2012).

Irrespective of the view of language adopted, authorities tasked with the design of language examinations such as the Grade 12 HL papers should ensure that the format is supportive of the construct that has been conceptualised, and that task types and items are aligned accordingly. In terms of the current study, the conceptualised construct of a generic and differential language ability resonates more closely with a view of language as a unitary construct than a sum of separate skills, which would suggest a need for the innovative design of integrated items that assess competence in a range of skills as well as media. This aspect will be addressed more fully in Chapter 5 as part of the analysis of the different examination papers and sections.

The ultimate goal of the study is to present a theoretical framework that would enable greater equivalence of standard between the respective HL examination papers. However, there can be little mention of equivalence without a clear understanding of *what* the language papers measure and *how*. An analysis of content and tasks is necessary to ascertain which abilities and items feature prominently in the papers and the desirability of continuing with the existing format of the papers. A close analysis of content should shed some light on what is being attended to in the language classroom and whether important components of language ability are being neglected. In order to find a way to introduce greater equivalence of standard between the different language papers, an articulation of both constructs (abilities) and levels of proficiency is needed. There would be little sense in developing HL papers that may be said to be equivalent in terms of construct, but not level of proficiency. Similarly, ensuring comparative levels while measuring entirely different constructs would prove to be just as unsatisfactory, if not impossible.

At this point the term construct needs clarification. In language testing literature, construct is associated with a number of other terms such as "blueprint", "rubric", "specification" and "trait" (see Lumley & McNamara 1999: 31; Davidson & Lynch 2002: 3; Davies, Brown, Elder, Hill, Hughes 2003: 26; Van Dyk & Weideman 2004a: 1; Weir 2005: 6), but treating these synonymously can only confuse matters. Although Davidson and Lynch (2002) prefer the term "specification", construct seems to be the word most widely used to refer to the overall ability or trait being measured. For the sake of clarity, the term "specification" will be used in this study to refer to the articulation of the construct in the detailed descriptions of examination tasks and items. The list of sub-abilities and accompanying task items to be performed in order to generate the needed evidence of the superordinate construct together constitute the blueprint of the examination. "Rubric" will be reserved for instructions on the marking and rating side of assessment.

1.3 Rationale for studying the HL examination

A definite bias towards English FAL (i.e. second language) over other school language subjects is discernible in the publications released by the Department of Basic Education and Umalusi. It would seem that the HL subjects are not accorded the same status or measure of interest by the education authorities. The annual report issued by the Department of Basic Education since 2011 to provide detailed feedback on the Grade 12 examination results does not seem to consider the HLs to be amongst the "most popular" or "key" school subjects (Department of Basic Education 2012c: 5). Of the school language subjects, only English FAL features in the report, which is aimed at improving learner performance.

The same prejudicial treatment of English FAL is to be found in the investigations and studies commissioned by Umalusi.[1] This council carries the responsibility of overseeing the quality of educational assessments under its jurisdiction, which includes the school examination system. Since the establishment in 2001 of the statutory body, which forms part of the National Qualifications Framework (NQF), considerable time and resources

[1] Umalusi is a name of Nguni origin that refers to a guardian of assets (Umalusi 2012c).

have been invested in quality assurance and research studies related to improving the standards of the curricula and the respective examination papers. As part of this initiative the National Senior Certificate (NSC) was introduced in 2008 to replace the Senior Certificate (SC) with its provincially set examinations, and a common exit examination that would be set nationally was implemented. Despite these efforts there are still discrepancies in standards between the various language examinations and the sets of scores obtained for these are not comparable across all languages and years, even though the language papers are based on the same subject assessment guidelines. For example, on average learners who offer English and Afrikaans at HL level score lower marks than those who offer other languages at this level. Varying degrees of difficulty and levels of cognitive demand in the examination papers have been cited as reasons for some of the disparities, as well as uncertainty about whether the same constructs are being measured (Umalusi 2012a). If a measure of equitability is to be achieved between the HL subjects, the clarification of constructs and standards is essential.

Viewed holistically, the study seeks to contribute towards the improvement of educational practice in the South African schools system. It has largely been inspired by the challenge of educational linguists such as Paola Uccelli and Catherine Snow to be "practice-relevant by design" (Uccelli & Snow 2010: 628) and provide information that may be of material use for the educators and their students in the classroom, the ministry of education and the local education authorities. Considering that South Africa still displays a number of inequalities in the sphere of education, after more than two decades as a democracy, addressing the issue of educational and assessment practices of varying quality and standards should be both of relevance and beneficence to the country.

The study also illustrates the importance of ensuring that assessment is attuned to standards and that there is reciprocity between the two. However, there is a real possibility that educators will understand the alignment of teaching and testing from the perspective of *teaching for testing*, a potentially harmful narrowing of subject content for the sake of achieving higher examination scores. Educators who are less experienced and poorly equipped to teach may thus resort to employing past examination papers as an abridged form of the language curriculum, rather than attempting to cover the prescribed syllabus. One possible way to improve the standard of the teaching in the classroom and prevent any narrowing of content would be to raise the standard of the examination assessment

by making the HL papers less predictable and ensuring that they cover a broad scope of the curriculum. Uccelli and Snow (2010: 638) emphasise the challenge to assess the more advanced language skills and proficiency needed in the higher grades, as "in the accountability-driven world of education, developing assessments for these more sophisticated language skills is key, because if they are not assessed, they are unlikely to be attended to in the classroom".

1.4 Scope of the study

This study falls within the domain of the discipline known as applied linguistics. Delineating the field of reference of this discipline, however, continues to be an elusive and contentious matter. At the one extreme scholars have argued the modernist case for a theoretical continuity in terms of which applied linguistics is regarded as a subdivision of linguistics. Towards the middle of the spectrum others have reconceptualised applied linguistics as a problem-solving enterprise and mediator between linguistics and other disciplines. The resultant contradiction that applied linguistics can both constitute an inherent part of linguistics, while at the same time falling on the continuum between linguistics and other disciplines, has yielded an alternative, postmodernist view, which lies towards the opposite end of the spectrum. It is a view that emancipates applied linguistics from the control of linguistic theory and acknowledges it as a discipline in its own right (Sealey & Carter 2004; Weideman 2007; Hall, Smith & Wicaksono 2011). All of these views, however, have had a significant role to play in attempting to define applied linguistics and in endeavouring to provide a theoretical foundation for language solutions to specific problems, particularly within the context of language development and education.

The attempt to define applied linguistics as a discipline of design is relevant since the latter is a reflection of a theoretical belief as to how language is learned or acquired, and, in the case of the present research study, more specifically how language knowledge and ability are assessed.

1.5 Thesis statement and research objectives

The focus of the study falls on two major questions originating in the examination section of the language curriculum for South African schools: What constructs are being assessed, and how can greater equivalence of standard be achieved between the respective Grade 12 HL papers, in order to prevent the prejudicial treatment of certain students? It is hypothesised that the HL papers are not comparable to one another because of a lack of consensus on the cognitive levels of challenge, confusion on which traits or sub-abilities should be measured and non-adherence to essential theoretical principles in language assessment. The latter would suggest that inadequate processes are in place to ensure comparability of standard and that the education authorities responsible for overseeing the HL examinations need to provide more comprehensive guidelines for the setting of the papers. It is further hypothesised that some components of the examination papers are weighted too heavily and not aligned with the designated notional hours of teaching and learning. The redesign of the format of the exit-level examination may therefore be necessary. As a result of these perceived inconsistencies, the validity and reliability of the HL papers may be questioned, as well as the fairness of using inferences based on examination scores as a basis for granting matriculants admission to institutions of higher education or access to work and employment opportunities.

The main research objectives, therefore, will be to achieve conceptual clarity on the superordinate construct and sub-abilities that should be assessed in the exit-level examination papers; determine what kinds of examination tasks are likely to generate the best evidence of language ability; and introduce a form of structural equivalence through the restructuring of the papers and application of uniform methods of scoring. The possibility of including an examination component that will be common to all the HL subjects as part of the restructuring process and as a viable means of attaining greater equivalence of standard will also be pursued.

1.6 Research methodology

A mixed methods approach will be adopted. The study has both conceptual and empirical components and incorporates primary and secondary research methods. After presenting the historical context behind the current system of educational assessment and identifying

existing disparities, a theoretical discussion of language testing principles is provided and a framework devised for the validation of the HL papers. Hereafter a description is given of the school HL curriculum and conceptual clarity sought on the superordinate construct and sub-abilities that are to be reflected in the corresponding examination papers. The latter is supplemented by primary, empirical research of a qualitative nature in which a content analysis is undertaken of a selection of examination papers to ascertain the extent of the alignment between the curriculum and examination papers. Task types and individual items are evaluated on the basis of the discussed language testing principles in order to determine their suitability as evidence-generating artefacts of language ability. As part of the analysis of the examination papers, a classification scheme is employed to generate quantitative data that can be used to identify trends and typicalities. A limitation of the study is that attention will solely be devoted to analysing the language component of the Grade 12 exit-level examination, i.e. Papers 1 and 3. The literature component (Paper 2) constitutes a distinct subject content area that warrants separate scrutiny and falls beyond the scope of the present research project. Nonetheless, cursory reference will be made to the weighting of Paper 2 as well as Paper 4 (oral school-based assessment) as part of the investigation into revising the format of the examination papers.

A further limitation of the study is that it will not be possible to supplement the qualitative aspect with quantitative statistical data of a detailed nature. The absence of raw scores for the respective examination items and sub-components rules out the possibility of determining the technical quality of the examination papers using either classical test or item response theory (Bachman 2004; Read 2010). The latter are particularly useful to show the reliability of individual examination items, and can facilitate equating different versions of examination papers. Although Umalusi has mentioned the desirability of investigating the use of item banking for the purposes of generating examination papers in the future, there is little evidence so far of any movement in this direction. Without the availability of the raw data required for psychometric purposes, no reliability or inferential statistics can be generated. It would thus not be possible to compare the performance of groups of learners writing different HL papers across different years of examination, or to identify areas of strength or weakness in language ability. Only the overall average percentages obtained per HL group are available from Umalusi. This means that no analysis can be made of how the respective examination items function (in terms of item difficulty and discrimination indexes), the consistency of measurement or

the correlation between different subtests and examination papers. Instead, evaluative and possibly subjective judgements may have to be made on the suitability of the task types and examination items. This is highly problematic, since determining the reliability of a measuring instrument is considered to be an essential part of the validation of a language examination (Jones 2012: 350).

As part of the validation process, the corresponding marking memoranda of the selected examination papers will also be subjected to evaluation in order to determine whether the system of score allocation supports the principles of reliability and validity. Again, the reliability and validity of the scoring need to be backed by empirical evidence through the statistical correlation of scores allocated by different markers, as well as the correlation of scores awarded by the same marker, i.e. intra- and inter-rater reliability indexes (Weir 2005). Attempts to access a sample of examination scripts in order to investigate the correlation of scores allocated by different markers were unsuccessful. Ideally, the qualitative and quantitative analyses should not be seen as dichotomous aspects, but as interactive and complementary (Van Dyk 2010: 21) and it would be preferable to be able to undertake both kinds.

Using the comparative data obtained from the content analysis, an attempt is made to increase the perceived validity and reliability of the examination papers by designing an alternative format for them. At the same time structural equivalence is sought which may provide a more equitable basis for comparing scores across different HL groups.

1.7 Research sample

In the light of the fact that there has been no substantial amendment to the format of the respective language papers since the introduction of the NSC in 2008, the analysis of a five-year sample of papers and their accompanying memoranda (2008-2012) is considered to be sufficient for the purposes of the current study. The intention is to scrutinise the English HL papers in detail, with occasional reference to the comparative sections of a selection of Afrikaans and Sotho[2] papers and memoranda. Since the researcher is based in the Free State province, the three dominant HL subjects used in the

[2] The English writing convention for the names of the indigenous African languages will be used.

schools of this region were selected. Use is made of translated texts to analyse the Sotho papers, as the researcher is not conversant with this language. Some meaning may be lost in the translation process, but it should still be possible to determine the main abilities being assessed and how marks are allocated.

The purpose of the comparative study is to illustrate how (in theory) Grade 12 language examinations can differ from one language to another in terms of what is assessed and the way marks are allocated for tasks, aspects which can contribute towards creating an unfair basis for contrasting performance between different examinees. The decision not to analyse a greater number of examination papers once again relates to the fact that this would serve little purpose. To date the HL papers have followed a similar structure to that of the English language paper, but there is no guarantee that the format and task types will not change in the future. There is thus little sense in investing considerable time in analysing further papers in detail when they could change at any point. Instead, the lessons to be gained from the detailed analysis of the English papers and the selection of Afrikaans and Sotho papers will be used to make recommendations for all the HLs. By drawing on universally accepted principles in language assessment, a generic theoretical framework can be devised that can improve the validity and reliability of all of the language papers. In this way, greater equivalence of standard can also be attained.

The decision to focus on the English papers also rests on the fact that the researcher is a teacher of English and has a greater knowledge of and interest in this language than any of the other official school languages on offer. Moreover, English has acquired the status of an international language, which means that the standard of teaching and examination in South Africa cannot be viewed in isolation from that elsewhere. A third reason would be the fact that English is the main language of instruction at tertiary level in the country, and that the school HL curriculum has as one of its core objectives the preparation of learners to cope with the demands of advanced language ability required at college and university. Fourthly, and most importantly, the principles of language assessment apply across all languages. Any theoretical framework reflecting a responsible approach to the design of language examinations would be suitable for employment by all persons tasked with setting the respective language papers, regardless of which language.

1.8 Exposition of the study

The Curriculum and Assessment Policy Statement (CAPS) introduced by the Department of Basic Education (2011a) to be fully operational by 2014, makes reference to communicative language teaching (CLT) and a text-based approach, but no analysis of the language abilities reflected in the policy document has been undertaken to establish the appropriateness thereof and whether the curriculum makes sufficient provision for the dominant fields of discourse (Halliday 1978, Foley 1990; Weideman 2009b) relevant to language learning at school level. The way abilities are assessed in the language papers also needs to be analysed to establish whether the kinds of language tasks typically included can provide adequate evidence of the differentiated and generic language ability that the curriculum refers to. Using the data obtained from the analyses of CAPS and the language papers it will be possible to validate the format of the English language examination and attempt to develop a framework for achieving construct validity and greater equivalence of standard between the different papers, as well as comparability across years of assessment.

Before proceeding to analyse the selection of examination papers, a sufficient understanding of the historical context of the official languages of South Africa and the dynamics of multiculturalism is essential. If a comparative basis is to be laid for the respective language papers, cognisance of the disparities that exist between the development and status of the indigenous languages is necessary. Chapter 2 is devoted to delineating this context and discussing how the dispensation of language assessment has changed since South Africa has become a democracy. Some of the current inequalities in the school examination system emanate from colonial ideologies and views from the apartheid era that continue to find a foothold in education practice to the detriment of many language learners at school.

The role played by the Council for Quality Assurance in General and Further Education and Training (Umalusi) is expounded and an overview is given of research commissioned by this statutory body relating to the school language subjects in particular. Already here it is evident that languages in South Africa continue to be treated differently and that we still have a long way to go if we are serious about acknowledging the importance of all

languages and giving learners an equal opportunity to succeed. An overview of research into the standards of the language curriculum is provided and the main findings discussed.

Chapter 3 includes a literature study of applied linguistics as a discipline of design and how developments in language testing have paralleled changing paradigms in the discipline. Historical trends in language testing are alluded to before proceeding with a discussion of important theoretical principles in language assessment. Particular attention is devoted to the validation frameworks of Bachman and Palmer (1996), Weir (2005), Weideman (2009a) and Green (2014). The main emphasis here falls on the notions of validity, reliability, practicality and fairness in language assessment and how these form part of a theoretical framework for the responsible and principled assessment of language ability.

A literature study in Chapter 4 informs the attempt to achieve conceptual clarity on what constructs and sub-abilities receive prominence in the curriculum and what the language papers are supposed to measure. Of importance in this section are the aims of CAPS (Department of Basic Education 2011a), theories of communicative language competence and socially informed ideas about language and language teaching (Halliday 1978; Blanton 1994; Weideman 2002; Larsen-Freeman & Anderson 2011). These serve as a basis for the formulation of a superordinate construct for the assessment of language ability at this level. Hereafter a content analysis of examination tasks and items in the Grade 12 English HL papers over the period 2008-2012 follows in Chapter 5. The objective hereof is to determine whether the examination papers display the essential qualities referred to in the framework proposed in Chapter 3. Hereafter, a comparative study is undertaken in Chapter 6 of structural elements in a selection of Afrikaans, English and Sotho papers to determine areas of commonality and to investigate different options to establish greater equivalence of standard and construct between the respective HL papers.

The design of an alternative format for the examination papers, the incorporation of new task types and the inclusion of a component common to all the language examinations, possibly as a separate paper, are explored in Chapter 7. Insights gained from the study are shared in the final chapter on the value of the research (Chapter 8). Recommendations are made that may be of benefit to education authorities, examiners, educators and students,

and that could potentially help to provide an impetus for the further development of all the HL subjects in keeping with the requirements of the Constitution (Act no. 108, Republic of South Africa 1996c) and the Language-in-Education Policy of the Department of Education (1997).

Clarification of certain concepts within the context of the study

Bantu language

The nine official indigenous African languages will be referred to as Bantu languages since the Department of Basic Education recognises Afrikaans as an "African language", necessitating the use of the linguistically correct term "Bantu language" to make a distinction where a reference to Afrikaans is not intended (Mesthrie 2002: 3-5; Department of Basic Education 2013).

Equivalence

This should be understood as a potentially subjective term aimed at communicating the extent to which two or more objects of study can be deemed to be related and comparable to one another in terms of sameness or similarity of standard, function, form and complexity (cf. Arffman 2010: 38).

Home Language

This refers to the highest level of language ability that is taught and assessed at Grade 12 level, and not necessarily the first language of the learners. The term "HL" does not convey the same meaning as that used in the Language-in-Education Policy where it is used to denote the literal sense also referred to as first language, or language used in the residential context of the family.

Validation

The term validation is used to refer to an investigative process to provide a rationale for the use of an applied linguistic artefact such as a language examination or curriculum. It involves a study into the qualities of the mentioned artefact and the systematic collection of evidence from a potentially wide range of sources for its validity and reliability on the

basis of a theoretical framework. The notions of validity and reliability are of fundamental importance to any validation framework and will be discussed in full in Chapter 3.

Assessment

This will be understood as the general term used to refer to the process of designing and administering different instruments and procedures to quantify knowledge of and ability in language. A language examination may then be considered as one form of assessment.

2 Historical context to the teaching and assessment of the HL subjects

2.1 Introduction

This chapter provides a historical and contextual discussion of where the current difficulties in the assessment of the HLs derive from. Despite several attempts to establish a more equitable educational dispensation since the transition to democracy, major differences are still discernible in the standard of the teaching and assessment of school subjects. It is true that some of these can be ascribed to socio-economic inequalities of the past that have yet to be eradicated in contemporary South Africa. It is a regrettable but sobering fact that learners who have the option to attend well-resourced schools will continue to be privileged with better facilities and teaching technology than those consigned to schools with less or even minimal resources. It is equally true that the provision of equal facilities and resources to all schools is logistically and practically unlikely to be achievable soon. Where there have been gains, the efforts are laudable, but it is certain that much still needs to be done.

Probably one of the biggest challenges that remains to be addressed is the varying qualifications and capabilities of educators and the extent to which this has the potential to compromise the standard of teaching (Bhorat & Oosthuizen 2008; Modisaotsile 2012). To compound matters further, the HL subjects that are the focus of this enquiry have not shared the same historical status and have not developed to the same extent (Louw 2004; Kamper 2006; Alexander 2013b; Webb 2013). There are, for example, different traditions of language teaching and testing among the languages taught as HLs, and learners of languages with strong oral traditions may not have access to as many written resources as those studying strongly developed languages such as English and Afrikaans (D'Oliveira 2003; Ministry of Education 2003). All of these aspects have a profound effect on the teaching and learning of languages, and by implication the assessment of language ability.

The first part of this chapter is devoted to a discussion of two key legal and policy considerations that are at the heart of this enquiry. First, there is the impact of the Constitution of South Africa, which speaks to the issue of fairness, and second, the

Language-in-Education Policy on the status of the school language subjects, that concerns the further working out of the constitutionally guaranteed equality among languages. Some factors that continue to constrain the development of the official languages in the school system are identified, as well as recent advances in the sphere of higher education that, despite the setbacks that have also been recorded, could serve as an impetus for the further development of these languages. This is followed by an overview of initiatives taken by the educational authorities to introduce a more equitable education system. In particular, the role of the Council for Quality Assurance in General and Further Education and Training, referred to as Umalusi, is expounded.

2.2 Parity of esteem and the move towards multilingualism

Nearly two decades have passed since the adoption of the Constitution of South Africa, Act No. 108 of 1996 (Republic of South Africa 1996c), and its recognition of eleven official languages. Since all the official languages are supposed to be accorded equal status, their equitable use, including the development of these languages as languages of education, is to be advanced through legislation, as evident in the following clause from the Constitution:

> (4) The national government and provincial governments, by legislative and other measures, must regulate and monitor their use of official languages. Without detracting from the provisions of subsection (2), all official languages must enjoy parity of esteem and must be treated equitably. (Constitution of the Republic of South Africa, 1996c)

One of the areas in which this advocated form of equality is to be strongly noticeable is the school classroom. Prior to 1994 English and Afrikaans, which were spoken by a minority of learners, had been entrenched as national languages and made compulsory at government schools for all learners to study, either as first or second languages. The Bantu languages, on the other hand, which were spoken by a great majority of learners, were marginalised in high-function domains such as parliament, printed media and institutions of teaching and learning where they were not used as the media of instruction (Mda 2000: 157-159; Webb 2013: 175).[3]

[3] Du Plessis (2003) provides a detailed summary of political and policy developments in language in education prior to and including the 1994 year of transition to a new democracy, in which the privileged positions of English and Dutch, and subsequently Afrikaans, are reflected.

Presumably because of the role that Afrikaans was considered to have played in the discriminatory policies and practices of the apartheid government and the fact that English was considered by many to be the *de facto* and liberating *lingua franca*, there was much debate by the newly elected government on whether English should be the only official language. However, owing to the calls of sociolinguists and educators for "cultural and linguistic pluralism", the nine Bantu languages were accorded official status and recognised as being equal to Afrikaans and English (Mesthrie 2002: 22). Clearly this decision was important for the sake of a peaceful transition to the new political dispensation, but it did signal a commitment to multilingualism and paved the way for the more equitable treatment (and by extension, development) of the languages as school subjects. Unfortunately, as this study will demonstrate, there are still a number of unresolved issues that continue to hinder the realisation of this objective.

2.2.1 Factors constraining the development of the official languages in the school system

One of the assumptions at the time of the new political dispensation may have been that entrenching equality for all the languages through the provisions of the Constitution and other forms of legislation, such as the Language-in-Education Policy (Department of Education 1997), would also guarantee the development and protection of all these languages. In this respect the two legal instruments referred to have helped to ensure the elevation and use of the eleven languages. Their success to promote these languages has, however, been varied, and they have not achieved the desired development or protection of the official languages in many spheres (Balfour 2006; De Kadt 2006; Webb 2013). English and Afrikaans are still the dominant languages of learning and teaching after the foundation phase at school, as well as at tertiary level. This means that first language speakers of these two languages receive plentiful opportunity to develop their academic literacy levels and proficiency in these languages during their years of schooling compared to speakers of indigenous African/Bantu languages.

The fact that the respective languages have had neither the opportunity to develop nor the resources to fully differentiate to the same extent as the dominant languages has implications for the teaching and assessment of these languages as school subjects. English is recognised as an international language with a rich heritage of literature

spanning centuries. Finding suitable texts and resources for educational purposes is easy. Although fewer texts are available in Afrikaans, a strong data base of literature can be accessed and there are sufficient materials available for the classroom. The situation with the Bantu languages is completely different, especially in the instance of a language such as Ndebele, which has relatively few speakers. On average only around 3300 learners write the Ndebele first language (HL) examination each year, as compared to the 110 000 who write the Zulu examination (Umalusi 2012a: 8). Such a disproportionate share of Grade 12 learners also has effects on the political influence of the respective languages.

It seems that texts tend to be created in some of the Bantu languages for use in the education system rather than for the purposes of public consumption (Umalusi 2012a). In other words, there are not many authentic texts to draw on for all eleven HL subjects. This may have implications for the constructs assessed in the language examinations and the types of tasks included, as well as the focus of teaching in the respective language classes.

It is also argued in the Umalusi report on the standard of the HL examinations (Umalusi 2012a) that because English and Afrikaans are the only languages of learning and teaching in Grade 12, and by implication the only languages in which other school subjects are examined, the amount of exposure at school that learners receive to these two languages is substantial. English and Afrikaans first language speakers thus have the opportunity to develop an academic register to a far greater extent than do their Bantu language speaking peers who receive exposure to their first/HLs for only 4.5 hours a week, during language classes (Umalusi 2012a: 7). However, this statement is somewhat contestable. It ignores the fact that Bantu language speakers do receive considerable exposure to their first languages outside the school context on a daily basis, but not necessarily in a written form or formal register. It is thus not accurate either to state or to assume, as that report does, that language classes are the only exposure that the learners have to their first languages. Printed media in these languages may not always be readily available or accessible to the learners, but there are radio and television stations that broadcast in all the official languages, and these would cover an array of genres and registers (see Du Plessis 2006 for a review of multilingual broadcasting and language policy).

Spoken discourse should also not be considered inferior to written discourse, as both modes are needed for knowledge acquisition. Halliday (2007: 95) points out that the "semiotic distance between home and school" needs reappraising, because the packaging of knowledge in "two very different ways, largely insulated one from the other" is unnecessary. He is of the understanding that spoken language, even when informal, is just as systematic as written language used formally:

> Spoken discourse is highly systematic; and the way commonsense knowledge is represented in speech is no less meaningful than the way any other kind of knowledge is represented in writing. When we look carefully – and linguistically – at children's real learning experiences, we find that there is clear register-type variation of the kind we mentioned, with these two typical formations or packagings of knowledge: casual speech, and formal writing. But at the same time, the child is learning through many different registers, spoken and written, all at once. There are no registers that are **not** used for learning. (Halliday 2007: 95)

Another disparity on the exposure side, used as an explanation for the difference in standards and referred to in the same Umalusi report on the HL examination papers, is that academic meta-language may be a problem for teachers of Bantu languages. It is stated that this is not a natural part of Bantu language discourse and "the context is therefore not as supportive for developing the kind of critical and close reading skills typically associated with the English examinations" (Umalusi 2012a: 7). Even if this were true, since it proceeds from the highly contestable assumption of some languages being (inherently) deficient, to what extent this can be considered a valid reason for not developing critical thinking and analytical ability in the Bantu language classroom is equally debatable. Also, a lack of knowledge of technical terms should not be allowed to interfere with the development of critical thinking skills. It is far less important to know and have at one's disposal all the specific terminology than to be able to ask probing questions relating to the content of a text or speech. Inferential comprehension can surely not be part of the English and Afrikaans curricula (or languages) alone. This would be contrary to the objectives and principles of CAPS, as the following extracts show (Department of Basic Education 2011a: 4-5):

> (c) The National Curriculum Statement Grades R-12 is based on the following principles:
>
> - Social transformation: ensuring that the educational imbalances of the past are redressed, and that equal educational opportunities are provided for all sections of the population;
> - Active and critical learning: encouraging an active and critical approach to learning, rather than rote and uncritical learning of given truths;

- High knowledge and high skills: the minimum standards of knowledge and skills to be achieved at each grade are specified and set high, achievable standards in all subjects;

(d) The National Curriculum Statement Grades R-12 aims to produce learners that are able to:

- identify and solve problems and make decisions using critical and creative thinking;
- work effectively as individuals and with others as members of a team;
- organise and manage themselves and their activities responsibly and effectively;
- collect, analyse, organise and critically evaluate information;
- communicate effectively using visual, symbolic and/or language skills in various modes;
- use science and technology effectively and critically showing responsibility towards the environment and the health of others; and
- demonstrate an understanding of the world as a set of related systems by recognising that problem solving contexts do not exist in isolation.

To imply that these principles and abilities are available only to some languages and not to others is to draw a line through the requirements set out in CAPS. Admittedly, varying levels of training and ability of educators do, however, play a role, apart from language traditions and the resources that are available. In this respect the teachers (and examiners) may be reluctant to change their styles of teaching and assessing, and in the process could be obstructing the intellectual development of the learners. Weideman, Tesfamariam and Shaalukeni (2003) have identified a reluctance amongst educators in general to adapt to new methodologies and imperatives. The implication of retaining approaches to teaching and testing that currently do not engage the learners in critical and creative thinking is that a greater investment will have to be made in improving the qualifications of the language teaching staff and providing some alternative forms of assistance to increase the likelihood of being able to offer instruction in higher order skills in the classroom. Without these, the equitable development of the school language subjects becomes less probable.

A brief review of how the Afrikaans language developed as a high profile language in South Africa suggests that much more can be done to elevate the Bantu languages, and that the school education system should play a pivotal role in increasing their spheres of influence and standard of use.

2.2.2 The development of Afrikaans as a language of academe

English and Afrikaans were not always of equal status in South Africa. Of the many languages that had been in use in South Africa by the start of the twentieth century, English and Dutch were the first to receive equal status in parliament and the judiciary when the South Africa Act of 1909 was passed (De Kadt 2006: 47). This act also heralded the start of an officially bilingual state. Already in 1902 English had been proclaimed the sole language of education in government schools. Up to then Dutch and Afrikaans had been used in many parts of the country, as had other indigenous African languages, but not on any official level. Afrikaans slowly gained more prominence and started to be used as an alternative to Dutch in schools from 1914 onwards. In 1920 Afrikaans was unofficially used as the language of teaching at the Universities of Stellenbosch and Potchefstroom, if on a limited scale. When in 1925 official status was accorded to Afrikaans shortly after the formation of a coalition between the South African Party and National Party, the idea of dual-medium education incorporating English and Afrikaans was introduced and bilingualism was made compulsory in the public sector (De Kadt 2006: 47).

The two languages were not equally developed at that stage. With its strong ties to Britain and other colonial territories, English already dominated the world of business, while Afrikaans was relatively undeveloped. Interestingly, many Afrikaans parents in the 1920s and 1930s objected to having their children educated through the medium of Afrikaans, believing that this would cause continued subjugation and place them in an inferior position to the English – the same paradigmatic kind of thinking that seems to be evident amongst some today with regard to educating children in a Bantu language, and ironically in respect of an education through Afrikaans too. De Kadt (2006: 49) draws a number of parallels between the situation that Afrikaans and the nine indigenous Bantu languages have found themselves in historically, which can be summarised as follows:

Table 2.1: Parallels between Afrikaans and the indigenous Bantu languages

Afrikaans prior to 1925	Indigenous Bantu languages prior to 1993
Fairly widely spoken as mother-tongue or HL	Fairly widely spoken as mother-tongue or HL
Extremely limited economic and political roles	Extremely limited economic and political roles
Speakers mostly poor, poorly educated, religious and rural	Speakers mostly poor, poorly educated, religious and rural
Used as language of education at primary level, but distrusted	Used as language of education at primary level, but distrusted
Fairly small literature	Fairly small literature

De Kadt (2006: 40) describes how within the span of a few decades Afrikaans developed from a language "which had no governmental recognition, and existed largely in spoken form" to a language which dominated considerable components of government, the economy and tertiary sector by the middle of the twentieth century, refuting the fears of parents that their children would be disadvantaged by an Afrikaans education. According to De Kadt (2006: 49), much credit is owed to the cultural organisations and formerly Dutch universities who played their part in helping to standardise and develop the language, in addition to the government's role in developing terminology. It seems that rather than waiting for the time when there would be a demand for Afrikaans in the business and other sectors before establishing it as a language of education, the latter became a strong impetus to develop the language. It was not long before Afrikaans graduates entered the labour market and made their language heard in the private and public sectors, increasing the prominence of the language and influencing its status. Today Afrikaans is still a prominent language in the economic sector.

After the transition to democracy, fears developed that Afrikaans would lose its strong position. Amidst speculations of a diminishing status of the language in the new South Africa, a process was initiated in 2003 to implement a language strategy that would promote and maintain Afrikaans as a "language of high function" (Prinsloo 2006: 138). A number of cultural organisations were involved in setting up this initiative, once again emphasising that help is needed from the language community to support the development and protection of a language. Speakers of the previously unrecognised

official languages could learn, it was thought, from the history and development of Afrikaans how education can be used to strengthen a language. Such strengthening does not need to wait for an appeal from the economic sector before proceeding with a strategy. This initiative found the lack of visibility of cultural and other organisations actively involved in the development of the Bantu languages puzzling. Could it be that speakers of languages with predominantly oral traditions do not experience the same need to extend the functions of these languages to written domains, as in the case of languages with strong written traditions? In other words, the fact that these languages are spoken by so many people in the home and social environment, as well as the informal sector, is deemed to be sufficient to guarantee their continued existence and use in these sectors. Afrikaans, on the other hand, has a strong written tradition.

De Kadt (2006: 50) believes the core reason for the difference between the development of Afrikaans in the pre-1994 era and the lack of development of the other nine indigenous African languages post-1996, relates to the nature of language recognition and how this varied between the two language groups. Afrikaans received its recognition when politicians representing the speakers of this language were elected to govern in an era when language and ethnic identity were the important political distinctions. It was the vulnerable position that the Afrikaners found themselves in with regard to the English that was "central to the construction of a strong Afrikaner identity, the key element of which was language" (De Kadt 2006: 49). In the case of the 1994 election, it was not one particular language or ethnic identity that was the distinguishing factor, but race and socio-economic class. The point she makes is that, although language is still important to the members of the new government, their being in power does not depend on their recognising and developing the Bantu languages. Given their historical opponents' strategy to use language both to mobilise ethnicity and to divide ethnically, it is perhaps understandable that they wish to take the opposite route: not to use language for the mobilisation of political power.

This brief periodisation of Afrikaans as a language of intellect and economic force nevertheless serves to illustrate how a language can be developed and its influence in society expanded through the school classroom. The point of view adopted in the current study is one that ascribes to using the HL subjects as a stronger forum for increasing their usability in important spheres, at least on a regional basis. Phrased differently, given the

constitutional imperative of ensuring the equitable treatment of languages, it would be unwise not to employ education as a primary means of nurturing and ensuring this. What is more, it is difficult to find an alternative explanation for the explicit goals of HL instruction at school in the official documentation and, specifically, in CAPS.

2.2.3 Future prospects for the development of the indigenous languages

The aim of Section 2.2.2 was to show that with sufficient prioritisation and investment unequally resourced languages can become more prominent and assume a more influential role in society over a period of time, even if they are unlikely to be in a position to become international languages such as English. In addition hereto, it is evident from the historical overview that the educational use of languages at school level can provide a valuable means of protecting and developing languages (De Kadt 2006).

Besides the educational sphere, Kamper (2006) believes that the indigenous languages have a strong role to play in community development and indigenous knowledge systems, but that (misguided) socio-cultural beliefs about their usefulness and a general lack of institutional support have had the opposite effect and contributed towards subtractive bilingualism rather than additive bilingualism. Rather than maintaining the first language (L1/HL), in addition to learning a second language, the development of knowledge and ability in the L1/HL has been neglected owing to the emphasis accorded to English as FAL and medium of instruction. Whether developments such as the passing of the Use of the Official Languages Act (2012) and the decision by the Department of Basic Education to introduce a third compulsory language subject in the school classroom (*The Times*, 7 August 2013) will be able to revitalise the importance of the official languages and promote increased bilingualism and multilingualism, remains to be seen. Perceptions that the Bantu languages are of little use other than in the home and school classroom (and there only partially) can change more easily if their use in high status spheres is increased (Alexander 2013a, 2013b; Webb 2013).

The higher education domain could be used to a greater extent to enhance the status of the Bantu languages. Up to now, however, a serious concern has been the dwindling numbers of students registering for Bantu language courses at South African universities (Kaschula 2013), and the realisation that any closure of Bantu language departments

would further be destructive to the existing indigenous knowledge systems and run counter to the ideals of multilingualism. Fortunately there are a few encouraging developments in the higher education arena, which may impact positively on some of the indigenous languages at least.[4]

The University of KwaZulu Natal recently announced its intention to introduce Zulu as a compulsory course for undergraduates and has already introduced Zulu language learning programmes for students of Nursing and Psychology (Kaschula 2013). In the northern part of the country, the University of Limpopo has already piloted the implementation of Pedi as a medium of instruction in certain degree courses (Kaschula 2013), while in the Eastern Cape at the University of Fort Hare, conversational Xhosa classes are being offered to lecturers who are not able to speak the language. At this institution all tutors employed at the university are required to speak Xhosa in order to facilitate the explanation of concepts in the first/HL of the students (Tshotsho 2013). At the University of Cape Town, the language skills of medical students are evaluated on site during clinical examinations in which they are expected to assist patients in an indigenous language. In order to graduate at this institution, medical students need to obtain credits for courses in Xhosa and Afrikaans.

These initiatives are indicative of a gradual move towards a more multilingual higher education environment that could serve to elevate the status of some of the HL subjects. A logical outcome would be to prepare learners to be fluent in academic discourse in more than one language at school level already. After all, preparing learners for tertiary study is one of the stated objectives of the HL curriculum. The idea of engendering academic discourse in the Bantu languages at school level to prepare them for university brings to the fore the important link between school and higher education. On this point, it is insightful to note the comment of Alexander (2013) on what may be interpreted as the tertiary sector's complicity (up to now) in constraining the development of the Bantu languages:

[4] For more detailed information on the intellectualisation of the indigenous languages, see Department of Higher Education and Training. 2012. *Green Paper for Post-Secondary School Education and Training.* Pretoria, RSA: Government Printers, and Department of Higher Education and Training. 2011. *Report commissioned by the Minister of Higher Education and Training for the Charter for Humanities and Social Sciences*. Pretoria, RSA. Government Printers.

> Above all other things, we have to remember the backwash effects of the university on the hidden curriculum in the school system. There is no point at all in valorising African languages at school, as we are now at last beginning to do, if this is negated by the language practices and attitudes prevailing at the universities. (Alexander 2013a: 83-84)

Irrespective of what may evolve at the level of higher education, the Constitutional prerogative remains to attend to the advancement of the official languages in as many spheres as possible. To achieve this at secondary school level, attention would need to be given to the development of an academic discourse and meta-language in the HL subjects, and ensuring that the prescribed materials and textbooks reflect developments in the higher education sector by containing suitable texts and terminology.[5] It goes without saying that all students should have access to materials and textbooks.

Although it may be argued that insufficient progress has been made to ensure the development and equitable use of all the official languages, credit needs to be given for what has been accomplished since the demise of the previous political dispensation some twenty years ago. This will form the focus of the ensuing section.

2.3 Post-1994 changes aimed at creating a more equitable system

The proclamation of eleven official languages can be seen as a first step towards introducing parity of esteem between the different language communities of South Africa and laying the foundation for access to equitable and quality education in the official languages. In as much as the democratic government sought to acknowledge diversity by increasing the number of official languages to eleven, on the education front the main objective shortly after assuming power was to consolidate structures by reducing the number of education departments and creating a unified system of education. The amalgamation of as many as 18 disparate education departments into one national department, the increased spending on education and redistribution of funds to more poorly resourced schools, and the formation of quality control organisations such as the National Qualifications Framework (NQF), have been hailed as amongst the most noteworthy achievements of the new government on the education front (Jansen & Taylor

[5] A number of universities are developing terminology lists in the local languages to assist students in the sciences. Some of this terminology can easily be incorporated into the school curriculum.

2003). Despite these achievements, the challenge to provide quality schooling to all students remains.

The proposed advancement of the Bantu languages at both tertiary and school level referred to in Section 2.2.3 can help to counter what Louw (2004) refers to as the Anglicisation of post-1994 South Africa. It is a sad fact that up to now English has retained its historically hegemonic position and is increasingly used in all sectors. In the school arena the number of Afrikaans-medium schools has dropped sharply with the prevailing perception (fuelled by government policies and practices and the globalisation of English) that, to succeed in South Africa, proficiency in English is needed above that in any other language (Louw 2004; Postma & Postma 2011; Webb 2013). The perceived bias towards English can be seen as a continuation of the initial campaign for English as the only official language advocated by many in the ANC leadership prior to the adoption of the country's Constitution. It is thus not surprising to note the tendency today of more affluent parents from diverse cultural groups (in the urban areas in particular) to enrol their children at English-medium schools where their first languages are not offered at all, making the policy of additive bilingualism in the schools difficult to implement. The policy advocates that learners should continue to learn their "home" (i.e. first) languages, while learning one or more additional languages (Department of Education 1997). This is not necessarily problematic, as middle class children do not as a rule suffer as a result of having a different language than their HL being used as medium of instruction at school. The real challenge lies in the larger numbers of children of working class parents in poorer, less well-resourced schools.

Notwithstanding the anomalies that exist, Umalusi records show that by far most learners are able to offer their first languages as HL subjects (Umalusi 2012a: 8), which is encouraging. Table 2.1 provides a four-year summary of the average number of learners writing the final Grade 12 HL examination.

Table 2.2: Average number of students writing each HL examination annually by province (2008-2011) (Umalusi Certification Database, Umalusi 2012a: 8)

	Eastern Cape	Free State	Gauteng	KwaZulu-Natal	Limpopo	Mpuma-langa	North-West	Northern Cape	Western Cape	Total
Zulu	4	681	14700	85825	752	8132	4		2	**110,100**
English	7095	2903	29200	25044	1940	3319	2391	952	14941	**87,785**
Xhosa	42596	628	3040	2055	3	48	384	442	8283	**57,479**
Pedi		8	6825		43045	5270	3			**55,151**
Afrikaans	4153	3492	12154	1027	1394	2490	2932	4500	17290	**49,432**
Tswana		1564	8774		468	1062	20037	3075	6	**34,986**
Sotho	972	15696	7724	30		149	553	17	64	**25,205**
Tsonga		20	1749		12194	6777				**20,740**
Venda			614		13725	2				**14,341**
Swati			18	12		13770	1	1	2	**13,804**
Ndebele			23		37	3265				**3,325**

Of the 472 348 learners depicted in the table, more than 80% offered a language other than English at HL level. Of the almost 20% of the learners who wrote English at HL level, approximately half are likely to be additional-language speakers of English on the basis of the latest census figures, which indicate that only about 9.6% of the South African population are first-language speakers of English (Statistics South Africa 2012).

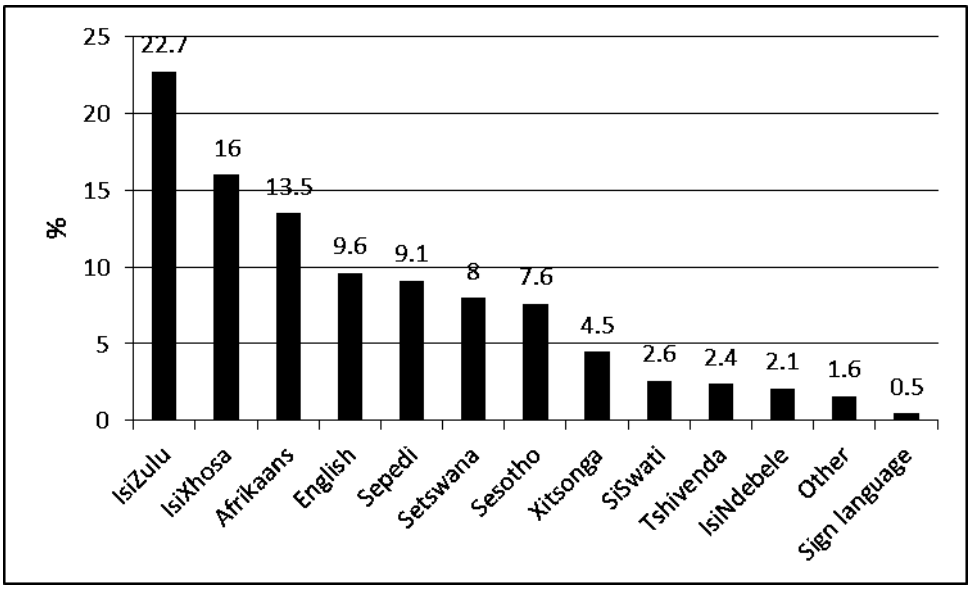


Figure 2.1: Distribution of the population according to percentage of first language speakers (Statistics South Africa 2012: 24)

It is regrettable that there are some learners who are unable to offer their first languages as a school HL subject. These are likely to be learners studying at the former model C schools in predominantly suburban areas that do not offer the Bantu languages at HL level referred to earlier (see Lafon 2008). This having been said, what is the primary focus of the present study is not how many schools offer Bantu languages at HL level, or whether learners are being deprived of the opportunity to study their first languages, but the standard of these language subjects and the extent to which the curriculum and exit-level examination papers reflect the desired outcomes. Up to now there has been little impetus to measure or benchmark the standard. In as much as the Umalusi report (Umalusi 2012a) mentions the possibility that some of the language papers appear too easy, the fact remains that there is no statistical basis for making such an assumption. There is a need for far more extensive research on each of the HL subjects to shed light on the issue of standards.

For the purposes of clarity, it should be pointed out the term "standard" in an educational context bears more than one meaning. In its plural form it refers to the detailed articulation of knowledge of subject content and ability required of students. In the singular, the word standard can be understood as referring to the level of achievement of the set standards. There are many questions surrounding the levels of teaching and assessment, but the Department of Education (now referred to as the Department of Basic Education) has devoted considerable time to the development of standards and made several changes to the curriculum (with mixed success) since 1994. These will be discussed next.

2.3.1 Post-1994 changes to the curriculum

In the apartheid dispensation there were separate syllabuses and examinations for school subjects. This is one of the reasons cited for the varying practices still discernable in the teaching and assessment of the language subjects:

> …different communities of practice have evolved with different assumptions, not only about standards, but also about the purposes of language teaching and assessment. (Umalusi 2012a: 12)

Although the Bantu languages used a common syllabus prior to 1989, shortly before the transition to democracy separate syllabuses were developed for most of these languages (Umalusi 2012a: 13). Standardising the languages was a core issue at the time and in fact remains a controversial matter (Webb 2008). While much corpus development has taken

place, acceptance of the standardised languages seems to be problematic. Also, educators are not necessarily communicatively competent in the standardised varieties since there are rural and urban varieties in use. Webb (2008) emphasises the role of educators, especially the HL teachers, in assisting with the standardisation process as role-models:

> As the appropriate variety in high-function formal contexts, in particular as a written form, the standard language must be accepted as the proper and appropriate target in first language study, and the aim of first-language teachers must be the development of learners' skills in using the standard variety in as many types of high-function formal contexts as possible ... Learners' linguistic skills in the standard variety of their primary languages must also be developed in content subjects, i.e. across-the-curriculum. (Webb 2008: 14-15)

Despite the fact that the Pan South African Language Board (PanSALB) established Provincial Language Committees (PLCs) for each province, as well as National Language Bodies (NLBs) and a number of lexicography development units in order to assist with the development of the indigenous languages, Webb (2008) argues that such efforts can only succeed in conjunction with strong role models such as educators, language practitioners and the media.

The efforts of PanSALB and the respective language bodies appear, furthermore, to have had little effect on the standardisation of school language teaching. Apart from the late standardisation of the indigenous languages by the language bodies (with the input of university language departments), different teaching methodologies have applied in schools. It is averred that greater emphasis has been accorded to structuralism and the formal teaching of grammar in the Afrikaans and Bantu language classrooms, while the English curriculum, which has drawn heavily on the British system, has devoted more attention to aspects of critical literacy (Umalusi 2012a).

The introduction of a common outcomes based curriculum in 1997 for all the official languages, the Senior Certificate (SC) (NATED 550 curricula; see Fiske & Ladd 2004; Umalusi 2012a), was an attempt to eliminate some of the disparities and provide common standards for language teaching and assessment. Further attempts were made to strengthen the school curriculum in 2001 with the introduction of the Revised National Curriculum Statement (RNCS) in terms of which all languages would henceforth be offered on both Standard and Higher Grade. Greater emphasis was once again placed on communicative competence, and learners were expected to study the same number of

prescribed works in each language. At this stage examination papers were set provincially. Besides the common curriculum and assessment standards, school-based continuous assessment (CASS) was introduced in a further attempt to establish equity and balance in the assessment division, as opposed to relying solely on the results which learners obtained in exit-level examinations. However, huge discrepancies have been reported between the marks awarded as part of CASS and the summative Grade 12 examination marks (Mncwago 2015). It was not long before it became apparent that much more revision was needed and the SC was replaced in 2008 by the outcomes-based National Senior Certificate (NSC) exit-level qualification with its curriculum comprising the National Curriculum Statement (NCS) for the Further Education and Training (FET) school phase (Umalusi 2010: 5). For the first time a common Grade 12 examination was set nationally, although the literature paper continued to be set provincially until 2009 (Umalusi 2012a: 13). A further notable change was the abolition of separate Standard and Higher Grade subjects. Future examination papers would have to include tasks that would distinguish higher achieving students from those performing on lower levels. To enable this, the Subject Assessment Guidelines (SAGs) specified the cognitive abilities that were to be assessed on different levels and the kinds of questions that should be asked.

From the Umalusi report (Umalusi 2012a), however, it can be inferred that the above interventions did not ensure equitable assessment of language ability across language papers. The report admits that one of the likely reasons for "the excessively high average marks in some Bantu languages is that the examinations are too easy" (Umalusi 2012a: 11). The average performance of students per HL group, calculated over a four-year period (2008-2011) and represented in tabular form below, shows a variation of up to 10% between some of the language subjects.

Table 2.3: Four-year average learner performance in the HL examinations: 2008-2011 (Umalusi 2012a: 9)

	Eastern Cape	Free State	Gauteng	KwaZulu-Natal	Limpopo	Mpuma-langa	North-West	Northern Cape	Western Cape	National Average
Sotho	61.2	54.8	57.0	56.6		57.8	54.6	55.1	56.2	**52.9**
English	52.1	50.4	55.0	58.1	49.3	49.5	51.6	55.7	54.2	**53.0**
Afrikaans	51.8	58.4	61.8	61.7	63.7	63.2	64.3	50.1	54.8	**55.5**
Tswana		54.8	55.6		56.2	63.6	59.3	55.0	62.1	**55.7**
Pedi		56.5	52.7		60.7	61.9	48.5			**57.4**
Xhosa	60.7	60.3	57.6				57.8	58.3	60.4	**59.5**
Zulu	68.1	49.4	60.1	63.7	58.5	51.6	51.9		55.0	**60.0**
Swati			64.4	67.7		85.1	33.0	53.0	59.0	**61.7**
Tsonga		58.7	53.8		66.2	64.3				**62.7**
Venda			56.7		65.4	53.3				**63.4**
Ndebele			58.5		65.3	66.9				**64.4**

An alarming discrepancy is the extraordinarily high average of just over 85% for Swati achieved by students in the Mpumalanga province and the corresponding disturbingly low 33% obtained by learners in the North-West province for the same HL subject. Such conflicting results demand close scrutiny and serve to illustrate that there are not only severe disparities in performance across languages, but within the same language examination.

The variation in averages is not the only point of concern. Compared to non-language school subjects, the average percentage of students who pass the HL subjects is exceptionally high (Umalusi 2012d: 61; Department of Basic Education 2012c: 5; Solidarity Research Institute 2015: 21-22). The following table reflects the average pass rates obtained per HL over the four-year period 2009-2012.

Table 2.4: Average national NSC pass rate per HL (2009-2012)[6]

	2009	2010	2011	2012	4 year average
English	93.2	92.8	94.0	95.3	94.0
Afrikaans	94.5	97.2	98.1	98.3	96.5
Sotho	97.5	99.0	99.3	99.7	98.8
Zulu	98.6	99.1	99.4	99.4	99.0
Pedi	98.5	99.3	99.1	99.6	99.0
Tsonga	98.9	99.1	99.3	99.2	99.0
Swati	98.8	99.2	99.4	99.3	99.1
Setswana	99.3	99.4	99.4	99.7	99.2
Ndebele	99.8	99.8	99.9	99.9	99.4
Xhosa	99.7	99.7	99.8	99.9	99.6
Venda	99.8	99.8	99.9	99.9	99.8
Average	**98.1**	**98.6**	**98.9**	**99.1**	**98.5**

The results reflected in Tables 2.3 and 2.4 may suggest that the standard of some of the HL papers is higher than that of the rest, but no such comparison is possible without a clarification of constructs and tasks. In fact, Umalusi (2012a) offers a tentative reason for the high averages in some of the language subjects as being the tendency to repeat literature questions, especially in the case of languages with few written texts such as Ndebele. Obviously, repetition would not only encourage rote learning, but also make it easier to answer the predictable questions. It is to be contested though whether there are so few literary texts in the minority languages to warrant the prescription of the same texts year in and year out, and even if the texts are used repeatedly, why the questions do not show enough variation across different years.

The exceptionally high pass rates for the language subjects are even more conspicuous when contrasted with the average pass rates in eleven other school subjects. The average obtained for English First Additional Language is commensurate with that of the HL subjects.

[6] Information compiled from data obtained from the Report on the National Senior Certificate Examination Results 2010 (Department of Basic Education 2010b) and the National Senior Certificate Examination Technical Report (Department of Basic Education 2012d).

Table 2.5: Average national NSC pass rate per key subject[7] 2009-2012 (Department of Basic Education 2012c: 5)

	2009	2010	2011	2012	4 year average
Mathematics	46.0	47.4	46.3	54.0	48.4
Physical Sciences	36.8	47.8	53.4	61.3	49.8
Accounting	61.5	62.8	61.6	65.6	62.9
Agricultural Sciences	51.7	62.6	71.3	73.7	64.8
Life Sciences	65.5	74.6	73.2	69.5	70.7
Economics	71.6	75.2	64.0	72.8	70.9
Geography	72.3	69.2	70.0	75.8	71.8
Business Studies	71.9	71.1	78.6	77.4	74.8
History	72.2	75.8	75.9	86	77.5
Mathematical Literacy	74.7	86.0	85.9	87.4	83.5
English First Additional Language	92.7	94.5	96.2	97.8	95.3

The observed variation in scores between different examination subjects is not unique to South Africa. In a study conducted by a team from Durham University in the United Kingdom (Coe, Searle, Barmby, Jones & Higgins 2008) in which a number of comparisons were made between the relative difficulty of different examination subjects on the basis of data captured from as far back as 1970, similar variations were recorded, although not to the same extent as in Table 2.5. The study confirmed that STEM subjects (Biology, Chemistry, Physics, Mathematics) are by far perceived to be the most difficult on average and that there are "substantial differences in the average grades achieved by the same (or comparable) candidates in examinations in different subjects" (Coe *et al.* 2008: 2). Nonetheless, an overly simplistic interpretation of what subjects are the most difficult on the basis of average scores is to be avoided. A number of other aspects could be responsible for the variation in scores, such as attitudes of examinees, motivation, time spent on teaching and learning and the quality of the teaching. As an alternative, Coe *et al.* (2008: 3) advise changing the way grades for subjects are used:

> This would involve introducing some kind of scaling so that some grades are acknowledged to be worth more than others for certain purposes. Existing statistical differences would continue, but a 'fair conversion rate' would be applied when grades in different subjects were to be treated as equivalent … (Coe *et al.* 2008: 3)

This is an aspect that warrants further attention in subsequent research studies, considering that the grades obtained in Grade 12 are used for university access purposes.

[7] Key subjects refer to those school subjects with the highest numbers of students, i.e. the most popular subjects (Department of Basic Education 2012c: 5).

At this point it can be stated that the average scores obtained for language subjects may vary considerably from those of the mentioned STEM subjects, but not to the degree reflected in Table 2.5. A variation of as much as 46% between some subjects is unacceptably high.

2.3.2 The role of Umalusi

The onus to find solutions to what historically has been a disparate and dysfunctional education system has fallen largely on the shoulders of Umalusi. This statutory body was established in 2001 by virtue of the General and Further Education and Training Quality Assurance Act No. 58 as amended in 2008, and is further governed by the National Qualifications Framework Act No. 67 of 2008. Umalusi forms part of the National Qualifications Framework (NQF) along with two other units, namely the Council on Higher Education (CHE) and the Quality Council for Trades and Occupations (QCTO). The NQF in turn falls under the supervision of the umbrella organisation known as the South African Qualifications Authority (SAQA) that is tasked with ensuring that qualifications meet the criteria prescribed by the NQF and that they are of acceptable quality and internationally comparable (Government Gazette Number 36797, Notice Number 648, p. 3, 30 August 2013).

One of the main tasks of Umalusi is to ensure the quality of educational assessments under its jurisdiction, which naturally includes the Grade 12 exit-level examinations. In this respect Umalusi is responsible mainly for the following (Umalusi 2013: 8):

- development and management of a sub-framework of qualifications
- quality assurance of assessment (both internal and external) at exit points
- certification

In terms of the above, Umalusi is expected to determine compliance with policy pertaining to examination related processes, as well as establishing the extent of cognitive challenge of examination papers, the weighting of the content in relation to the curriculum and the standard of the marking of examination scripts (Umalusi 2006: 2). It is the primary task of Umalusi to ensure that those organisations or institutions which are accredited to provide education and training in South Africa are delivering quality learning programmes and qualifications and that the accompanying assessment processes are

reputable. Commissioning research to ensure educational quality thus forms an integral part of the mandate.

Even before the introduction of the NSC and the outcomes-based curriculum, assessment practices were identified as an area of contention in the report on *"Making educational judgements: Reflections on judging standards of intended and examined curricula"* (Umalusi 2007, 2011b). The need to consider the role of technology as an aid to determining standards and improving the examination system was highlighted in the report of 2008 which covered the possibility of item banking and Item Response Theory (IRT) as a means of introducing equivalence between examination papers. Specific mention was made in this report to the necessity of "conceptual clarity of the underlying construct" as a requirement for employing any form of psychometric measurement (Umalusi 2011b: 23).

When accusations of a lowering of standards were levelled at the education authorities after the introduction of the Senior Certificate, Umalusi initiated a number of research projects to verify whether standards were in fact dropping. These can be summarised in terms of three main project categories: research reports published on prevailing standards, benchmarking exercises with foreign qualifications, and the holding of an international conference to debate issues of equivalence and quality. The matter of how to articulate standards has in fact become a central issue in the work and deliberations of Umalusi. Attention will be devoted to the main findings of the respective initiatives in the section that follows.

2.3.3 Research studies commissioned by Umalusi between 2003-2012

A number of research reports on curricula and standards have been released by Umalusi in the past decade. Research was conducted in four main strands as indicated in Table 2.6 (Umalusi 2011b: 9).

Table 2.6: Main strands of Umalusi research 2003-2012

Strand 1	Quality Assurance Methodology
Strand 2	Further Education and Training *Sub-strand 2.1 Senior Certificate/National Senior Certificate* *Sub-strand 2.2 Vocational college subjects at Senior Certificate Level*
Strand 3	Adult Education and Training *Sub-strand 3.1 Occupational qualifications* *Sub-strand 3.2 Adult Basic Education and Training (ABET)*
Strand 4	General Education and Training *Sub-strand 4.1 Foundation phase*

Only those reports pertaining to the language component of the Further Education and Training phase in Strands 1 and 2 will be discussed for the purposes of the current study.

2.3.3.1 Strand 1: Quality Assurance Methodology

"Approaches to quality assurance in the GET and FET bands" (Umalusi 2004a)

In this report, which was published in 2004, the need was identified for a more explicit system of quality assurance that incorporated the development of standards per qualification, as well as outcomes pertaining to curriculum content. At the same time it was acknowledged that the introduction of pre-specified standards and learning outcomes would not guarantee any improvement in the quality of the qualification.

The examination model of ensuring quality of education was re-affirmed as the most practicable form of assessment, provided there was a standardised syllabus to accompany the set standards and outcomes. It should be noted that this system has a strong washback[8] effect, in terms of which the questions set in the examination papers tend to determine what is formally taught in the classroom.

The quality of qualifications was to be assured predominantly through the monitoring of the standard of question papers and the standard of marking, in addition to the monitoring of what was covered in the curriculum and syllabus. Moreover, the examination system should afford students the opportunity to show their ability to apply and analyse learning material and their capability for critical thinking and creativity. The analysis of a selection

[8] This is generally defined as the effect that language testing has on actual language teaching and learning, as well as on educational practices, beliefs and systems (see Bachman & Palmer 1996: 30-31).

of question papers and memoranda later in the study will reveal to what extent this is being achieved.

In addition to the examination papers which would be subjected to external assessment, teachers would be responsible for the assessment of coursework. The latter would also undergo a process of external moderation before being used to determine the final grade mark of a student.

"Making educational judgements: Reflections on judging standards of intended and examined curricula" (Umalusi 2007)

The second report was published in 2007 and was the culmination of three other research projects spanning a four year period in which the South African Matric was scrutinised, schools and colleges were compared, and the curricula of countries such as Ghana, Kenya and Zambia were investigated. The main aim of the study was to draft a comprehensive and revised curriculum containing clear curriculum statements and outcomes to replace the Senior Certificate. Of particular relevance on the assessment side is the mention in the report of the necessity of a taxonomy or grid to evaluate cognitive challenge and higher order thinking in question papers, and the revision of marking memoranda.

"The role of IRT in selected examination systems" (Umalusi 2008a)

This study examined the role of technology in high-stakes testing contexts, specifically the advanced examination systems in the Netherlands and Western Australia (well-resourced countries) and Indonesia (a poorly resourced country using advanced testing technology). In these countries use is made of item banking (the electronic storage of a selection of test or examination questions and tasks which have been piloted and analysed) and Item Response Theory (IRT) is employed to ensure equivalence across different administrations of examinations. The provisional finding was that it would be advisable to use a combination of subjective assessment and objective or empirical measurement in the South African system. Moreover, a prerequisite for any empirical enterprise was that there be clarity on the constructs to be measured, an aspect that forms an integral part of the current study and that relates directly to the important but hitherto neglected issue of construct validity.

The Umalusi report recommended that the role of IRT and its practical implications be investigated further. Not much use has been made in South Africa of IRT, and in particular the Rasch model. So far it would seem that mainly classical test theory has been employed for the purposes of statistical analyses of large-scale testing such as the academic literacy testing administered under the auspices of the Inter-institutional Centre for Language Development and Assessment (ICELDA), though some analyses reported on ("Research" tab, ICELDA 2015) do include differential item functioning (DIF) analyses on comparisons across different administrations of a test in which one parameter logistic modelling (OPLM) is used. Any introduction of item banking and IRT would inevitably require financial investment, a worthwhile venture considering that this would help to ensure more objectivity and greater reliability across different years of testing.

2.3.3.2 Strand 2: Further Education and Training

"Investigation into the standard of the Senior Certificate examination" (Umalusi 2004b)

The main objective of this study was to determine whether there had been a lowering of standards in some of the Grade 12 Senior Certificate examination subjects based on the NATED 550 curricula during the period 1992-2003. English First Language and English Second/Additional Language (as they were referred to at the time) formed part of the investigation. In particular the focus fell on three years: 1992 (when there were 19 different education departments responsible for administrating examinations); 1999 (when low pass rates were encountered); and 2003 (when pass rates were higher than ever). Syllabi, examination papers, memoranda and marked scripts were scrutinised. As far as the language component was concerned, the key findings were that there had been a decline in standard for the higher grade English Second/Additional language and English First Language examinations. The degree of conceptual challenge had also diminished. Other concerns were the quality of the marking and moderation, as well as the effect of the statistical moderation process on pass rates. It would seem that the CASS marks had played a part in inflating the 2003 pass rates. This study was followed up by a specific investigation into the quality of the school-based continuous assessment (CASS).

"Learning from Africa" (Umalusi 2008b)

The objective of this study was to reinforce the new South African curriculum and system of examination by comparing the local syllabi and examinations to those used in other

Anglophone African countries, specifically Ghana, Kenya and Zambia. Amongst the conclusions reached were that the outcomes focus of the South African curriculum made it difficult to distinguish between English HL and English FAL, as well as between different grades. Recommendations were made for the consideration of examination anchor items which would assist with the comparison of standards and working according to a system of grading marks rather than setting a pass mark. Grade boundaries could be determined per question paper, instead of per subject as a whole, which would enable further differentiation. Once again the importance of including a system of item banking is foregrounded in this report.

"Signalling performance: An analysis of continuous assessment and matriculation examination marks in South African schools" (Umalusi 2008c)

As part of this study, data obtained from CASS were evaluated against the results of the matriculation examination for English First Language and English Second Language over a span of three years (2003-2005). The data included the results of all Grade 12 students from the nine provinces. The externally moderated examination results were considered to be the most accurate assessment of the students' performance. Two aspects in particular were measured: the extent of the gap between CASS marks and examination marks; and the extent to which CASS marks correlated with examination marks. The findings of the study were that CASS tended to be a more lenient form of assessment leading to inflated marks, and that it was an unreliable form of assessment. Moreover, the results of the study had no impact on subsequent CASS marks, as teachers failed to amend their assessment practices, leading to a continuation of the mismatch between the CASS marks and external assessment standards.

"2008 Maintaining Standards Report" (Umalusi 2009a, 2009b, 2009c)

As part of this study a comparison was made between the curriculum and examination papers of the Senior Certificate (NATED 550) for the period 2005-2007 and those of the new National Curriculum Statement (NCS) for the year 2008. English First Additional Language was one of the six subjects investigated. The finding of the study was that English FAL in the new curriculum was more difficult than that of the NATED 550 curricula, and that the 2008 examination papers for this subject were closer to the Higher Grade SC papers than to the Standard Grade papers. Presumably in the light of the context the comparison was made with English Second Language, although this is not stipulated.

A limitation of the comparative study is the fact that, apart from the nationally set papers 1 and 3, only provincially set examination papers from the Western Cape and KwaZulu-Natal were used for the comparison of examination paper 2, which covers literature. Also, the syllabus documents from the NATED 550 curriculum were those used by the House of Representatives, an aspect which definitely undermines the reliability and generalisability of the research findings. Another shortcoming of the language leg of this study is to be found in the divergent ways the researchers used and interpreted the taxonomy used to determine which test items fell into what type of cognitive demand category and what level of difficulty they were. Although an attempt was made to align all of the different findings on the basis of an adapted taxonomy, the decision as to which category and level applied to which examination item remains of a most subjective nature.

"Evaluating the South African National Senior Certificate in relation to selected international qualifications" (Umalusi 2010)

A benchmarking study was requested by Higher Education South Africa (HESA) in 2008 to determine to what extent the NSC could be considered comparable to foreign qualifications. In collaboration with HESA, Umalusi benchmarked five subjects, including English First Additional Language, against the curricula and examinations of the following:

- Cambridge International Examinations (CIE)
- International Baccalaureate (IB)
- Namibian National Senior Secondary Certificate (NSSC) (considered to be a contextualised version of the Cambridge qualification for the Southern African region)

Both Cambridge and the International Baccalaureate are recognised as credible international education programmes of long standing. They are operational in more than 125 countries, including a number of African countries that share with South Africa common contextual challenges.

It is significant to note that whereas in South Africa the distinction between higher grade and standard grade was abolished soon after the advent of the new political dispensation in the country, the above-mentioned qualifications continue to differentiate between candidates of different levels, abilities and interests. The IB offers both Standard Level and Higher Level, while the CIE distinguishes between what is referred to as "the AS

Level and the A Level" (Umalusi 2010: 6). Cambridge also offers what is called a Pre-University curriculum (Pre-U), in addition to the A and AS levels in the advanced course component. It would seem that even though the A and AS levels are aimed at preparing learners for university study, there is a need for further differentiation in terms of learning content and needs through the Pre-U curriculum.

Part of the purpose of the comparative study with Cambridge and the IB was to assist Higher Education South Africa (HESA) to set admission requirements for candidates who had obtained international qualifications from these institutions and who wished to study at a South African university. Why the decision was made to compare English First Additional Language rather than English HL is unclear from the Umalusi report. Predictably, the research team found that the qualifications were not comparable owing to the divergent nature and foci of the curricula. The NSC, CIE A-level and the qualifications offered by the IB, were all deemed adequate for admission purposes to higher education (Umalusi 2010: 165). From this one can infer that in HESA's opinion English First Additional Language is sufficient for the purposes of preparing for a tertiary education in South Africa. However, research studies such as the Alternative Admissions Research Project (AARP) initiated at the University of Cape Town show that the standard of English FAL is not adequate for university purposes, and that students battle to "process academic reading and writing at the level required of an entry-level student" (Cliff & Hanslo 2009: 268). The need to address the standard of the HLs is further borne out by studies at a number of universities that it is not only FAL learners who battle with the language demands of higher education, but English and Afrikaans HL learners as well (Van Dyk, Van Dyk, Blanckenberg & Blanckenberg 2007).

"Comparative analysis of the National Senior Certificate HL Examinations, 2008-2011" (Umalusi 2011c)

This is the first study commissioned by Umalusi on the standard of the HL subjects. The emphasis here fell on adherence to subject assessment guidelines, degrees of difficulty of examination items, cognitive demands and suitability of the standard, quality, language and format of the language papers. Two main limitations were identified by the research team, namely that the judgements of examination items by the panels used in the study may not have been balanced (each language team was comprised of only four members), and that different interpretations were possible for the cognitive categories used.

Furthermore, the analyses of the respective language papers revealed that allowing students to choose from a selection of task items in some papers made it difficult to judge to what extent the Subject Assessment Guidelines had been complied with in the design of the paper and that the option to select tasks had also enabled candidates to determine how demanding a question they would like to answer. The general finding was that weightings of cognitive demand varied across the different papers and among the respective languages.

Umalusi decided not to publish this report, but launched a follow-up study which will be discussed next.

"The standards of the National Senior Certificate HL Examinations: A comparison of South African official languages" (Umalusi 2012a)

As part of the Maintaining Standards project, this study sought to examine the degree of difficulty and cognitive demand of the respective HL subjects owing to concerns about high pass rates. Trends in examination results across languages over a four-year period (2008-2011) were studied, and teams of evaluators evaluated the cognitive challenge and level of difficulty of each HL examination using an adaptation of the taxonomy of Barrett. The main findings of the study were that some papers appeared to be too easy, but owing to the subjective nature of the evaluation, inconsistencies with the way the taxonomy was applied and the absence of statistical data, it was impossible to compare standards across languages. The English Paper 1 was deemed to lack sufficient lower level questions, while the other languages seemed to have too many in the lower order and too few questions of an inferential nature. In this sense the findings were similar to those of the previously mentioned unpublished report. Moreover, as in the case of the previous study it was found that the vast choices of questions available to examinees in Papers 2 and 3 made it difficult to compare cognitive challenge. Barrett's taxonomy was also found to be unsuitable as an evaluation instrument for all sections of the papers.

The teams of evaluators were concerned about three further matters. Firstly, apart from the English papers, grammar questions tended to be limited, far too easy and at times unrecognisable, especially in the case of the Bantu languages (possibly owing to standardisation issues). Secondly, some texts were considered biased and unsuitable. Lastly, the poor translation of texts used for some of the minority languages was a serious

concern. These aspects will be taken into consideration when recommendations are made in a subsequent chapter and once the empirical analyses of a selection of examination papers have been reported.

"Developing a framework for assessing and comparing the cognitive challenge of HL examinations" (Umalusi 2012b)

With still no conclusive findings possible from the previous studies on the standard and cognitive challenge of the HL papers, Umalusi commissioned the development of a framework that would help to assess and compare cognitive challenge in the respective language papers. The latter needed to be able to accommodate students performing at widely divergent levels, which meant the inclusion of a vast range of questions requiring varying cognitive ability. A revised version of Bloom's taxonomy was adopted. This version, however, did not correspond fully with the taxonomy of Barrett used by the Department of Basic Education to evaluate cognitive levels of examination papers. The first part of the study recommended combining the taxonomies of Bloom and Barrett to reflect three levels of cognitive demand (low, medium and high) and that Umalusi find a way for the difficulty level of each question to be judged more inferentially. Further to this, the writing demands of responses should be taken into consideration, as well as the degrees of difficulty of written and graphic texts.

The second part of the research study was devoted to generating a framework of reference to assist evaluators to judge the levels of difficulty of examination questions, and the design of an instrument to compare the HL examinations in a more transparent way than previously. The report acknowledged that the judgement of question difficulty was compounded by the influence and interaction of diverse elements and that the evaluators varied in their ability to make connections between different examination items, an aspect that requires a high level of expertise. The proposed framework should rather be seen as providing a means for evaluators to reflect more deeply on each question on the basis of certain principles and to facilitate discussion during the pre-examination phase.

The proposed instrument for analysing the HL examination papers introduced a distinction between cognitive demand and level of difficulty of questions. In addition to the revised taxonomy incorporating Bloom and Barrett's previous taxonomies, a framework was provided for identifying sources of difficulty on the basis of the subject

content or conceptual knowledge being assessed, the difficulty of the stimulus (linguistic features of the question and texts), difficulty of the task (steps required to carry out the task or produce an answer), and the difficulty of the expected response format (as indicated in a mark scheme and memorandum). In as much as the above efforts should be commended, it is clear that the process of determining levels of ability and degrees of difficulty remains subjective. It is also unlikely that teams of evaluators would be able to reach consensus on each examination item, or that markers would approach responses in a uniform way and allocate marks consistently. In fact, the quality of the marking could undermine the best designed of examination papers, rendering the application of any framework redundant.

If all of the above strategies are taken into consideration, it is not surprising that one of the techniques Umalusi resorts to in order to ensure equivalence of standard across different administrations of examinations is that of statistical adjustment.

2.3.4 Statistical methods to achieve equivalence in the assessment of HLs

Umalusi is responsible for ensuring that the exit-level examination papers are comparable from one year to the next in terms of their standard and quality and that the distribution of results shows consistency across a span of time. In order to achieve these objectives, a system of moderation is applied which is referred to as the "standardisation" of marks (Umalusi 2011a: 3). The rationale expressed for the standardisation process is the need to "mitigate the effect of factors other than learners' knowledge and aptitude on their performance" (*ibid.:* 5).

The main assumption on which the principle of standardisation rests is that in the case of sufficiently large cohorts of examinees, there should not be much variance between the aptitude and intelligence of the learners across periods of time. The principle thus applies that if equivalence of standard is achieved between two or more examinations, certain statistical patterns should be evident which should correspond from year to year. According to Umalusi's system of statistical moderation, the distributions of the current year's examination results are compared with the corresponding average distributions of the results of a number of examinations from previous years. The raw marks of the

preceding three to five years are used to determine the "historical average" (Umalusi 2011a: 6) which is then compared to the average obtained in the current year. Any deviations in the distribution of scores require investigation and may be attributable to the distinctive nature of a group of candidates offering a particular subject, disruptions to the school programme and differences in the quality of teaching and learning provided at schools. If no "valid reasons" can be found to explain the non-conformity in the statistical distribution of results, "it should be accepted that the differences are due to deviations in the standards of the examination or of the marking, and the marks should be adjusted to compensate for these deviations" (Umalusi 2011a: 6).

If we consider the enormous differences to be found between the thousands of government schools pertaining to school management, training and qualifications of teachers and resources available (see Bhorat & Oosthuizen 2008), it is to be doubted whether parity is in fact achievable and whether the adjustment or standardisation of marks could compensate for the educational losses resulting from dysfunctional schools. It may be argued that only when schools have the same educational resources in all respects and learners the same opportunity to learn, can one begin to speak about achieving equivalence between standards and examination papers. Some global form of mark compensation may be possible in the interim within the existing quality assurance process, but certainly no redress is immediately available for students subjected to inferior teaching.

In addition to the challenge to ensure equivalence of the external assessment, the marks of the learners obtained through internal assessment are supposed to correspond with the adjusted examination marks that the same candidates achieved in the examination. This may necessitate the adjustment of the internal assessment mark in the absence of such correspondence.

A further indicator of mark correlation is to be found in the comparison of marks of related subjects. In this sense, candidates who perform well in one language paper, should also do well in a second or third, failing which there may be a difference in the standards of the respective language question papers. Should there be compelling evidence of disparities in standards and equivalence, Umalusi (2011a: 7) undertakes adjustments to the marks on the basis of the following guidelines:

a) If the distribution of the raw marks is below the historical average, the marks may be adjusted upwards, subject to the limitation that no adjustment should exceed half of the actual raw mark, i.e. half of what the candidate got, or 10% of the maximum marks for the subject.
b) If the distribution of the raw marks is above the historical average, the marks could be adjusted downwards, subject to the limitation in a), above.

Umalusi acknowledges that the standardisation of marks is not the ultimate solution to overcoming variations in standards, but at least it affords a partial means of achieving comparability and equality. In view of the problematic nature of ensuring equivalence in assessment, the current study proposes the incorporation of pretested and calibrated examination items to achieve greater parity of standards and increased credibility of the examination system.

2.3.5 The first international Umalusi conference

The first international Umalusi conference was held in Pretoria from 10-12 May 2012, with the theme "Standards in Education and Training: The Challenge", bringing together educationists and ideas from all over the world. One of the main caveats issued at the conference was that the setting of standards would not necessarily result in better or comparable education and that it would be unwise to rely on the new Curriculum and Assessment Policy Statement (CAPS) to remedy the situation. As Prof. Michael Young from the Institute of Education at the University of London pointed out in his keynote address, knowledge is not something that can be expressed in standards. Of more importance according to Young was what he referred to as "epistemic access" – the shaping and guiding of enquiry into truth.

Another concern raised at the Umalusi conference was the inadequate investment on the part of the education authorities in the development of teachers, an aspect which another keynote speaker, Prof. Catherine Snow of the Graduate School of Education at Harvard University, considered to be the real key to raising the bar. The role that language played in all subject fields was another major point that resonated throughout the conference, with a number of research studies demonstrating a link between language proficiency, reading ability and academic success. Closely related to the importance of language was the issue of literacy in the 21st century and how academic literacy development needed to be prioritised from the early grades. According to Prof. Snow in her keynote address, the definition of literacy had shifted in the 21st century and learners needed to read for

understanding and develop conceptual representations which could serve as a basis for the evaluation of new information. Such knowledge bases could be built in any language and from the youngest of ages. In the light hereof the diversity of languages in a multilingual country such as South Africa should not be viewed as an impediment to literacy development, necessitating the adoption of English as the only language of teaching and learning.

Another problematic area in education that was foregrounded at the conference related to misconceptions about the role of assessment, as well as varying standards in continuous or school-based assessment. It would seem that teachers in general struggle to interpret and use curricula to achieve assessment standards that are in congruence with those standards required by examination boards.

The current study intends to generate strategies to address some of the shortcomings identified at the Umalusi conference, particularly those pertaining to the development of academic literacy and the need to develop higher order thinking. The proposed redesign of the examination papers could serve as a vehicle in this regard.

2.4 Conclusion

From the historical overview it is evident that the adoption of the Constitution and Language-in-Education Policy created a platform for the development of the HL subjects, but that their status in education today is not comparable. Indeed, there are still conflicting views on how they should be taught and assessed. Developing a critical literacy through the HL subjects seems to be an issue in particular. Further to this, the extent to which educators are well equipped to facilitate language learning, the issue of what standardised varieties should be used for some of the Bantu languages and the unequal resourcing of schools are compounding factors. Nonetheless, there are some initiatives, such as those on the higher education front that were referred to in Section 2.2.3 above, that may increase the status and perceived usefulness of some of the Bantu language subjects and serve as a stimulus for prioritising their standard and ongoing development.

Much has been accomplished by the education authorities to introduce parity through the establishment of a national education department and quality control organisations such

as Umalusi, as well as the introduction of a common curriculum and a nationally set examination for all the language subjects. The concerted efforts to improve the curriculum (although with disappointing results) also deserve mention. It is obvious, however, that these interventions have not ensured comparability of standard, neither in respect of teaching and learning, nor in terms of the assessment of language ability. Attempts by Umalusi to determine the cognitive levels of ability reflected in the respective examination papers and the degree of difficulty of examination items were undermined by a lack of consensus amongst the members of the evaluation teams, and the absence of statistical analyses reflecting item difficulty and discrimination indexes. These are important aspects of classical test and item response theory that show statistically what items examinees find easy or challenging, and that also indicate the consistency of the measurement instrument (cf. Bachman 2004: Chapter 5). From the statistics generated on the basis of overall percentages obtained, far too great variations are discernible between some of the HL results, and the close to 100% pass rates cast doubt on the credibility of the results. The proposed framework developed by Umalusi for assessing and comparing the cognitive challenge of the language papers is a necessary step as part of the systemic validation process and in engendering greater consciousness of what is being assessed, but has shortcomings and is unlikely to ensure greater equivalence of standard. It may also be too sophisticated an intervention to be effectively implemented in our current education context. At this stage the framework serves merely as a *post hoc* point of reference and has not been adopted officially as part of the design process (personal communication with Umalusi). It is thus unlikely to have any effect on the kinds of questions that have been the feature of the HL papers up to now and also not likely to address the issue of covering the required levels of cognitive challenge, irrespective of what taxonomy of cognitive demand is recommended.

Of particular concern in this section of the study is the little consideration that has been given to the articulation of the actual constructs being assessed in the language papers and the principles underlying their design. Not only has their lack of articulation been identified in many reports commissioned and reported on here, but in the project brief by Umalusi for this investigation, it was also singled out as the most significant issue. In what follows Chapter 3 examines different theoretical frameworks for the validation of the HL papers, an essential step before proceeding with the detailed analysis of a selection of examination papers. As will be shown, understanding the theoretical principles of test

design and language assessment is crucial for identifying shortcomings in the existing papers and making appropriate recommendations.

3 A framework for the responsible and principled design of the HL papers

3.1 Introduction

The HL papers are high-stakes assessments with political, social and economic implications. Examination scores obtained by matriculants are used for employment purposes and for gaining admission to tertiary institutions, in many cases with the aid of bursaries or loans. Ensuring the credibility and usefulness of the HL papers is thus essential in light of the fact that the outcome of the examinations has the potential to create new forms of inequality and social injustice. In order to undertake an analysis of the HL examination papers, and in the interests of ensuring a measure of fairness and accountability in the examination system, a framework that is founded on theoretical principles and backed by empirical research is of paramount importance. Although a theoretical framework *per* se will not guarantee similarity of construct and equivalence of standard, it can provide a valuable reference point for comparing performance within and across different language groups. This will also facilitate alignment of the HL papers with current theories and applications that can contribute towards more equitable and socially responsible assessment of language ability.

This chapter will conclude with the articulation of a framework for responsible language assessment design which if employed may serve to facilitate the restoration of the credibility and trustworthiness of the HL papers. However, before discussing the essential qualities of language assessment artefacts and how these contribute towards the construction of an argument for responsible test design – a process conventionally referred to as validation – drawing mainly on the frameworks of Bachman and Palmer (1996, 2010), Weir (2005), Weideman (2009a) and Green (2014), a brief historical overview of developments in the field of applied linguistics and language testing is necessary to understand the theoretical context underlying the HL curriculum and assessment of language ability.

In the exposition below of a number of paradigms that have been influential in applied linguistics, a platform is offered from which we can subsequently analyse how these

alternating styles of working in the discipline affect the particular part of that discipline, language assessments and their design, in which this investigation has been conducted.

3.2 Traditions of applied linguistics

Language teaching and testing form part of the discipline of applied linguistics whose beginnings can be traced in particular to the audio-lingual era of language teaching that gained ground before and in the 1960s (Weideman 2017: 150). Through examining the historical development of the discipline from that time, insights can be obtained on the content of analysis and emerging trends in applied linguistics. Weideman (2017: 173 ff.) identifies seven phases in applied linguistics since the inception of the discipline, all distinct but relative and building on the previous. A study of the focus and main interest of the successive traditions discloses why applied linguistics cannot be defined as a mere extension or application of linguistics and why it should be viewed as a discipline in its own right.

Table 3.1: Seven successive traditions within applied linguistics (Weideman 2009b: 62, 2017: 174)

Paradigm/Tradition	Characterised by
Linguistic/behaviourist	"Scientific" approach
Linguistic "extended paradigm model"	Language is a social phenomenon
Multi-disciplinary model	Attention not only to language, but also to learning theory and pedagogy
Second language acquisition research	Experimental research into how languages are learned
Constructivism	Knowledge of a new language is interactively constructed
Postmodernism	Political relations in teaching; multiplicity of perspectives
A dynamic/complex systems approach	Language emergence organic and non-linear, through dynamic adaptation

The first generation of applied linguistics shows a strong belief in and reliance on the study of linguistics to inform language teaching and language learning. In this initial phase of the discipline an attempt was made to prescribe and validate language teaching techniques "scientifically" on the basis of linguistic and behaviourist theories deriving from the work of Skinner and Thorndike (Lado 1964). Not surprisingly, the first attempts to delimit applied linguistics were influenced to a great extent by prevailing Western

thinking that science provided a means to obtain sound knowledge and that the employment of technology was a form of applied science. Such essentially technocratic thinking culminated in views that technical-scientific methods could be used to analyse and engineer humans and society, which naturally included the lingual aspect of reality (Weideman 1987). As a result hereof, the first tradition discernible in applied linguistics placed considerable emphasis on some form of purported scientific proof. An example hereof is the application of behaviourist theory to models of language learning characteristic of the middle of the previous century. Weideman provides a strong critique of the modernist perspective that science alone can guarantee authoritative solutions to language problems and points out that scientific analysis itself is not neutral. Predictably, the supposed benefit of scientific analysis for applied linguistics characteristic of the first generation of applied linguistics has been rejected in postmodernism and has opened the door for alternative views that do not consider applied linguistics to be an extension of linguistics but a fully fledged discipline of its own.

The view that applied linguistics was an extension of linguistics derived partly from evidence of a bi-directional and reciprocal feedback between the two disciplines (Weideman 1987). Accordingly, applied linguistics was perceived to be the carrying over of linguistic knowledge into language teaching. The theoretical assumption was that language could be acquired through dissecting it into small structural units as linguists did and then teaching these fragments incrementally (Weideman 2017: 150). Corder (1973: 31) illustrates this perspective with a statement that a comprehensive plan for a language-teaching operation "must be expressed in 'linguistic' linguistic terms – lists of grammatical structures and vocabulary …" and that the linguistic approach determines "how we describe what we are to teach".

This view was contested with the arrival of transformational-generative grammar in the 1970s. In terms of this theory human beings possess an innate cognitive (mental) ability to learn and use languages and the rules that accompany them (Ouhalla 1999). When no evidence could be found of a mentalist approach in the prevalent language pedagogy and teaching materials at the time (Weideman 2007), scholars were faced with the predicament of explaining why linguistic theory was not being reflected in language teaching. Consequently, the notion of applied linguistics as a continuation of linguistic theory started to lose its firm footing and applied linguistics came to be seen as a

discipline in its own right (Sealey & Carter 2004[9]; Weideman 2007; Hall, Smith & Wicaksono 2011).

Apart from the inability of linguistic theory to explain language teaching methodologies in operation, the modernist view of applied linguistics advocated by some proponents of linguistic theory was also criticised for its positivist and prescriptive focus on a supposedly scientific foundation (Weideman 1987). On the matter of the monotonously repetitive audiolingual method of teaching sound systems and sentence patterns characteristic of the first tradition of applied linguistics, Weideman (2007: 591) states that rather than providing any demonstration of the application of linguistics to the design of a solution to a language problem, "the 'linguistic paradigm' of first generation applied linguistics … has left us with a language teaching design devoid of proper theoretical justification". In addition, transformational-generative grammar of the 1970s also failed to acknowledge the instrumental communicative function that language fulfilled. As a result hereof, linguistics lost some more of its grip on applied linguistics, which came to be seen instead as a mediating discipline. Weideman (2007) believes that, although the mediating perspective is problematic, developments such as the above have helped to emancipate applied linguistics from its direct dependency on linguistics as mother discipline.

The proposition that applied linguistics fulfils a mediating role remains troublesome, since, of necessity, in order for there to be a mediating role, the nature of the two things being mediated needs to be entirely different. Yet, if the one is considered to be part of the other, the implication is that the two are not inherently different. Rather, as Weideman (2017) shows, there is a difference in principle between the two, with applied linguistics operating in a much more specified and contextualised environment, a view shared by Sealey and Carter (2004). The study of language and linguistic concepts can therefore not be equated with the making and application of language plans, teaching courses and language examinations as instruments of design to address an identified problem. The two aspects may be related, but applied linguistics cannot simply be seen as a continuation

[9] Sealey and Carter regard applied linguistics as a social science and see language use as a form of social practice, hence their view that social science disciplines are better able to describe certain aspects of linguistic behaviour than are those disciplines which are concerned primarily with language. They redefine applied linguistics as "problem-based researching into communication-mediated issues in social life" (Sealey & Carter 2004: 17).

of linguistics, since the latter deals with an analysis of the learning and use of language and the structure of lingual objects, while the former attempts to address a language problem – and usually a large-scale or at least pervasive one – in a particular and complex context through the design of a solution. It is in view of its reciprocal relationship to language teaching that language assessment is considered to be a branch of applied linguistics. By describing test construction as "essentially a matter of problem solving, with every teaching situation setting a different testing problem", Hughes (2003: xi) and others who associate language testing with solving problems, indirectly support the view of applied linguistics as a discipline of design.

The difference in emphasis between linguistics and applied linguistics can also be discerned in the explanation of Widdowson (1984) that the term applied linguistics indicates the use of theoretical studies of language to generate solutions to problems arising in different domains, without assuming that a relevant model of language must of necessity derive from a formal model of linguistic description. A more recent view is that of Hall, Smith and Wicaksono (2011: 15) who employ the wording "autonomous applied linguistics" to emphasise that applied linguistics is not limited to any application of the findings of general linguistics. Their perspective corresponds with that of scholars such as Brumfit (1980) and Weideman (2017) that the scope and methodology of the subject fields differ, and that the real issue is the investigation of solutions to real-world problems in which language features as a central issue. Hall *et al.* define autonomous applied linguistics as a "discipline concerned with the role language and languages play in perceived problems of communication, social identity, education, health, economics, politics and justice, *and* in the development of ways to remediate or resolve these problems" (2011: 15). As such applied linguistics draws on multiple theories and methodologies from other fields, rendering the notion of "linguistics applied" redundant, an aspect that is made clear in the multi-disciplinary model advocated by third-generation applied linguistics.

The fact that theories started to be developed from work already done within applied linguistics is described by Weideman (2017) as the discipline's coming of age. The point to be noted in the search for a theoretically justifiable basis for applied linguistics as a discipline, however, is that "in designing solutions to language teaching problems, theory does not lead the way" (Weideman 2007: 594). Widdowson (1984: 8) in fact states that

the relevance of linguistics to language teaching cannot be assumed, and that it is likely that "linguistics, as customarily conceived, may *not* be the most suitable source for a practical teaching model of language". This point is illustrated by the failure of both behaviourism and cognitivism to provide any enduring theoretical basis for language learning and teaching, and by the move towards communicative language teaching (CLT) in the 1980s. It was only after the implementation of CLT designs and the adoption of an enriched view of language as social instrument in the second generation of applied linguistics that research on second language acquisition and constructivism, the focus of fourth and fifth generation applied linguistics, generated a theoretical justification for the designed methodologies and teaching practices already operational in the language classroom (Weideman 2017: 158).

The above discussion shows the continuity and reciprocity between the successive traditions of applied linguistics, and at the same time the innovative attempts in each phase to generate new solutions to language related problems. In the linguistic "extended paradigm" model of second generation applied linguistics we see an attempt to "redefine and extend the linguistic basis of the work done by the founding fathers of the discipline" (Weideman 2017: 151). There was a shift in language teaching away from a restrictive view of language, which emphasised teaching "formally defined units", to a functionally defined view of language that incorporated "larger, socially relevant units of language" (*ibid.*: 151). From this we can infer that it was the failure of linguistic theory to provide a methodology for meaningful language instruction that prompted a new design for language curricula at first devoid of theoretical (linguistic) backing.

Weideman (1987; 2011) explains the above phenomenon by referring to what he terms logico-analytical and technical-analytical analyses. In terms hereof, linguistics may be conceived as the insights gained through a theoretical analysis of the lingual mode of experience, whereas applied linguistics should be viewed as those insights obtained through an analysis of a language problem with the purpose of resolving the latter in a technically designed solution. Linguistic knowledge may thus subsequently be used to identify a language problem and justify a technical design that will provide the solution. The fact that the anticipation of a design is referred to suggests the dilemma of attempting to provide applied linguistics with a scientific status in terms of which a particular method of language teaching or assessment may be deemed to be fully scientific and henceforth

foolproof, acceptable and credible. The complex nature of the subject field implies that such a notion of scientific status is unrealistic and inappropriate. Since science is founded on theory and not absolute truth, as is evident in the evolutions of language teaching methodology and changing philosophical paradigms of teaching and testing, the inference to be drawn is that science and theory can never be neutral or fixed concepts. Any object of study will inevitably be subjected to the influence of political, social, cultural and other realities of a changing nature. The third generation multi-disciplinary model of applied linguistics in the middle of the 1980s illustrates the realisation that "solutions to language problems need a wider than linguistic justification" (Weideman 2017: 154) and that contributions from other fields of study such as psychology and pedagogy may be incorporated to address multi-faceted problems. This departure from relying solely on linguistics paved the way for empirical investigations into second language acquisition in the fourth tradition of applied linguistics and later approaches of a postmodernist nature in sixth generation applied linguistics. This shift away from linguistics as the mother discipline also resonates with dynamic systems theory in which language growth (and by implication its assessment) is seen as the "outcome of complex interactions among a multiplicity of systems, not all of which belong to the language organ constituted by the mind" (Weideman 2017: 166).

A further example of the reciprocity of different traditions can be seen in the CLT designs that developed from the second tradition of applied linguistics and that reflected elements of both modernist and postmodernist thinking. The humanistic emphasis in early CLT is considered by Weideman and others to foreshadow "critical, postmodern applied linguistics" (Weideman 2017: 153). Much contemporary thinking in postmodern applied linguistics derives from the consciousness of political power relations in language teaching and testing and the need for political and social accountability in relation to language solutions. Though this shift elucidates what alternative conceptualisations of applied linguistics might be entertained, it does not fundamentally alter the disciplinary character of the field. Weideman (2007) observes that although postmodernist approaches signal a break with their modernist predecessors, discontinuity is an impossibility, since the latter continue to define them, albeit negatively.

From the preceding overview, it is evident that each of the first six traditions of applied linguistics referred to has played a part in helping to define the subject field. Furthermore,

although each may be described as having distinct features, the connections between the six do not follow a strictly chronological order. Rather, they serve to challenge and redefine one another in an asynchronous fashion, re-affirming their ongoing relevance for future developments in the field of applied linguistics. The seventh tradition, deriving from complex systems theory, draws attention to the fact that aspects of previous schools of thought may re-emerge in later paradigms, whether in a similar or new format. It would thus be a mistake to attempt to base applied linguistics solely within any one particular tradition: "Each paradigm potentially offers a unique set of principles according to which applied linguists may design solutions to language problems" (Weideman 2017: 174).

Although a succinct definition of applied linguistics that satisfies all parties eludes us, at least the above understanding provides a philosophical framework within which to proceed. Taking into account the diverse aspects raised in the aforegoing discussion, the current study will be undertaken within the paradigm of applied linguistics as a discipline of design in which theoretically justifiable solutions are sought for complex language-related problems such as the challenge to design language assessments of a comparable construct and standard. The following section shows how alternating paradigms of applied linguistics have reflected in language testing practices over the past century, and how different task types that have become common features of language assessment instruments reveal philosophical views on what language is and how the ability to use it can be measured.

3.3 Shifting paradigms in language testing

The history of language testing as a practice that is self-reflective and theoretically founded is of a brief nature, although tests have been in use for many centuries (see Spolsky 1995; Weir, Vidaković & Galaczi 2013; Green 2014). Notably, the incorporation of language testing as a branch of applied linguistics assisted with the expansion of the testing profession. As a result of the influence of applied linguistics, various language testing theories developed in the field alongside the institutionalisation of language testing in the middle of the twentieth century on an interdisciplinary basis incorporating especially the field of psychometrics (Spolsky 1995; McNamara & Roever 2006).

The proposed theoretical framework for the purposes of evaluating the HL papers demands specific cognisance of past and current trends in language assessment. In his seminal work on the history of language testing, *Measured Words*, Spolsky (1995) identifies two dominant ideologies spanning the past roughly 100 years of testing in the United Kingdom and America (where institutionalised language testing has a strong tradition and history), and different design approaches, which at times have overlapped. The first ideology, which he terms the "humanistic-scepticist descriptive approach", corresponds with what he refers to as the "*pre-scientific*" or "*traditional*" format (1995: 5), while the second, which he refers to as psychometric-structuralist, relates to counter-attempts to measure language ability scientifically in a quest for certainty. The two ideologies continue to compete throughout the 20th and 21st centuries, with the British/European school preoccupied more with the trait being tested (i.e. construct), and the Americans concerned rather with how testing takes place and whether it is reliable (Weir 2005; Weir, Vidaković & Galaczi 2013). There is merit to be found in each of these and it is not surprising to see combinations of both ideologies featuring in new versions of existing tests (cf. Spolsky 1995).

The dominant phases in language testing spanning the past 100 years and how they relate broadly to traditions of applied linguistics and approaches to language teaching that were referred to above, are summarised in the table that follows.

Table 3.2: The influence of applied linguistics on formal language teaching and testing (based on Weideman 1987, 2002, 2007; Spolsky 1995; Weir 2005; Green 2014)

Applied linguistics tradition		Language teaching approach	Language testing approach	Typical assessment techniques and emphases
Pre-modernism		Grammar translation Direct method	Pre-scientific/traditional	Translation and grammar exercises Essay writing Open-ended interviews
Modernism	Linguistic/behaviourist	Audio-lingualism Separation of skills	Psychometric-structuralist	True/false items Multiple choice and discrete item tests of grammar, vocabulary, phonetic discrimination, reading and listening comprehension Focus on reliability
	Linguistic "extended"	Natural approach	Psycholinguistic-sociolinguistic	Cloze procedure, dictation
		Communicative/task-based approach	Communicative	Assessment tasks intended to reflect "real life" language use Tasks that integrate skills Focus on validity of content
	Multi-disciplinary	Communicative/task-based approach	Communicative	Assessment tasks intended to reflect "real life" language use Tasks that integrate skills Focus on validity of content
	Second language acquisition research	Communicative/task-based approach	Communicative	Assessment tasks intended to reflect "real life" language use Tasks that integrate skills Focus on validity of content
	Constructivism	Communicative/constructivism and socio-cultural approaches	Communicative/assessment for learning	Interactionist/supported forms of assessment such as portfolio work
Postmodernism/dynamic systems approach		Post-method/focused eclecticism	Critical language testing	Former techniques combined with new integrated kinds of tasks Focus on consequential validity/test impact Collaborative/interactional tasks

The table reveals a broad variety of testing techniques that at different stages in the history of language testing have all been considered (often uncritically and unjustifiably) to be credible and acceptable forms of assessing language ability. However, it is evident that merely having at one's disposal a selection of test techniques from which a choice can be made would be inadequate for the purposes of responsible test design. There is in the first place a need to understand the theoretical and philosophical views of language behind the selection of teaching methodologies and testing techniques, as the discussion on the discipline of applied linguistics in the preceding section has shown.

In the pre-scientific/traditional phase language was treated as an "object of study" in which the emphasis fell predominantly on appreciating the literature and culture of the language. Grammar-translation was used as a method by means of which grammar points were practised through the translation of sentences, in the belief that this would help the language learner to access the literature of the target language (Weir, Vidaković & Galaczi 2013: 16). This explains the continuing reliance on essay type questions and grammar exercises on the assessment side. Criticism of this method initiated a "new pedagogical approach rooted in the spoken language" (*ibid.*: 17) and an emphasis on phonetics and pronunciation – hence the inclusion of oral assessments to assess conversational fluency. In the HL papers we see strong examples of pre-scientific traditional forms of testing in Papers 3 (Writing) and 4 (Oral), as well as the emphasis on literature (Paper 2) especially. The grammar-translation method was specifically aimed at gaining access to the canonical literature of the target language – Virgil, Cicero in Latin; Shakespeare in English. Paper 1 (Language), however, reflects some elements of later applications of modernism, which is the first indication that contrasting approaches to language assessment are being used in the Grade 12 examination.

We can see that there was little regard for reliability of assessment in the traditional approach to language testing which was firmly entrenched by the end of the nineteenth century and which relied largely on the subjective intuition of expert examiners. However, already in 1888 Professor F.W. Edgeworth had alerted attention to the uncertainty and potentially erroneous scoring of assessments that required subjective marking, suggesting how these "could be made more precise by application of a part of

probability theory"[10] (Spolsky 1995: 23). The issues of subjectivity and unreliability in marking traditional examination papers also raised doubts about the fairness of the form of assessment, and paved the way for more objective testing in the first tradition of applied linguistics, which became a feature of modern language teaching from the 1960s on and which incorporated statistical qualities and technical forms of reliability, particularly in standardised tests. The issue in modern language testing was on accuracy of measurement rather than a concern with the social consequences of measurement and any potentially negative effect on test takers – aspects that would be foregrounded later in postmodern critical language assessment. There was also very little regard for beneficence towards test takers from different socio-economic and educational backgrounds in the traditional phase of testing, and test takers were expected to have access to extemporaneous sources of knowledge. Prinsloo (2004: 87) criticises this kind of language assessment which requires the production of examination essays as "reasoned social comment, a form of writing that assumes a middle-class location" (Prinsloo 2004: 87). The latter has been described by contemporary language testers as *ecologically insensitive* assessment that fails to interpret examination scores "in relation to a specific learning environment" (see Fulcher 2010: 2). In view of the inequities deriving from past educational traditions and practices alluded to in the previous chapter, examination authorities in South Africa need to guard against unfair and biased examination tasks that rely on extemporaneous knowledge and a privileged form of literacy.[11]

The move in the United States towards psychometric-structuralism as a "scientific" and objective means of assessment in line with the first tradition of applied linguistics signaled a radical departure from the traditional approach to language assessment with its characteristic essay questions. A problematic feature hereof, however, was the decontextualised assessment of knowledge and rules of structural grammar through discrete-point forms of testing. The latter derived from the influence of linguists such as Lado (1961) who believed language could best be taught by dividing it into linguistic elements and the separate skills of "listening, speaking, reading and writing" (Green

[10] For a discussion of how probability theory can be used in statistical analyses for language assessment see Bachman 2004: 224-226.

[11] It is clear from recent studies in the field of academic literacy development at tertiary level that there is still a preoccupation with "essayist literacy practices" (Boughey 2013: 28-29) and that little recognition has been given so far to other types of literacies that influence the ability to process information (see also Carstens 2012). Recognising multiple literacies could help to place students on a more equal footing.

2014: 179), which then needed to be assessed separately too. The purported advantage of discrete-point items was the opportunity to assess a wider variety of language knowledge which would (so it was believed) provide an indication of overall proficiency. Later traditions disputed this and showed a preference for task-based authentic forms of testing of an integrative nature which emphasised the communicative function of language in specific contexts over the expressive (see Alderson, Clapham & Wall 1995; Truscott 1996; Weideman 1987, 2002). Attempting to define language as a combination of sound, form and meaning (as studied in phonology, morphology, syntax and semantics), necessitating the assessment of separate language skills, was subsequently considered inadequate (Blanton 1994; Bachman & Palmer 1996; Van Dyk & Weideman 2004a). At the same time, the behaviouristically inspired and decontextualised discrete-point testing of components of language came under increasing criticism from applied linguists such as Corder and Selinker (Green 2014: 185). Notwithstanding this, discrete-point testing remains useful for some testing contexts and continues to be used for diagnostic purposes in postmodern language teaching (Hughes, 2003: 19), but not as an indication of overall proficiency.

Although the new objective task types such as multiple choice items introduced in the psychometric structuralist phase helped to provide consistency of measurement, language assessment should not be seen as merely the combination of language with existing forms of psychometric measurement. McNamara and Roever point out that "a psychometrically good test is not necessarily a socially good test" (McNamara & Roever 2006: 2). It would seem that since language is inextricably linked to a social context – an awareness that developed in second-generation applied linguistics – the social dimension of testing may be expected to be more marked in language testing than it is within the ambit of general cognitive ability assessment. In the opinion of McNamara and Roever, the cognitive bias in psychometrics has hindered some aspects of language testing and more research is needed on the social effects of testing, an aspect referred to as the "social turn" (McNamara 2005: 775) of language testing to which we will return later.

Despite the importance of consistency of measurement that was acknowledged in the psychometric-structuralist phase, dissatisfaction with attempts to advocate a scientific manner of teaching and testing characteristic of the 1960s and 1970s generated a move towards recognising psycholinguistic and sociolinguistic aspects of language learning,

and paying attention to the measurement of functionally and socially contextualised language use. Test developers realised the need to relate performance in a language test to the use of that language in a specific setting (Bachman & Palmer 1996, 2010; Weir 2005; Weir, Vidaković & Galaczi 2013). It then followed that if language performance was the outcome of language ability, this ability that was to be assessed needed to be defined before any test could be constructed.

Integrative task types such as cloze procedure and dictation became features of the linguistic "extended" phase of applied linguistics. Cloze procedure entails the mechanical or selective deletion of letters or words in a text and then requiring test takers to use contextual clues to reconstruct the omitted components. Proponents of cloze in the second generation of applied linguistics, such as Oller, claimed that it approximated actual language use and could measure integrated skills. Oller based his "unitary competence hypothesis" on the belief that there was a "single principal factor underlying all language skills" and as such cloze procedure was able to indicate "total language proficiency" (Weir, Vidaković & Galaczi 2013: 70). However, others criticised cloze as an indirect form of measurement for productive abilities such as writing. The main objection was that although it represented a more natural approach than discrete-point item types, it failed to provide sufficiently for the interactional nature of discourse. Nevertheless, cloze procedure has remained a popular item type in language teaching and testing, even though its effectiveness as a substitute for assessing productive ability remains contentious (Alderson, Clapham & Wall 1995: 44; Hughes 2003: 189; Weir, Vidaković & Galaczi 2013: 180). Where practical considerations preclude the direct assessment of productive ability, cloze procedure is still recommended in view of the fact that cloze procedure scores have shown a correlation with writing and speaking scores (Van Dyk & Weideman 2004b: 20; Weir, Vidaković & Galaczi 2013: 71-72). As such, it would be a useful examination task type for the assessment of HL ability.

In the same linguistic-extended phase of applied linguistics, dictation came to be viewed as a way of assessing listening skills, but lost popularity (probably because it is time consuming to administer and score) until studies showed high correlations between dictation test scores and the results of more complex tests (Hughes 2003: 195). The Pearson Test of English, Academic (PTE-A), which was introduced in 2010, includes dictation (Green 2014: 175) and we can expect this older task type to resurface in some

language classrooms. It could possibly be included as part of school-based assessment of HL ability.

The dominance of the communicative approach evident in Table 3.2 in the second to fifth generations of applied linguistics shows its pedagogical capital as a popular and effective way to learn to teach and to design the learning of language in an instructional setting. Although the preceding traditional and behaviourist traditions may have prepared the way for communicative language teaching and task-based assessment, this should not be seen as having provided a theoretical justification for the communicative approach prior to its implementation: "… the history of CLT provides a good example that an influential approach to language teaching can develop without initial theoretical justification" (Weideman 2017: 110). As mentioned earlier, the theoretical backing for CLT was provided *ex post facto* during fourth-generation applied linguistics research into second language acquisition.

The communicative approach that gained ground in the 1970s placed the emphasis on being able to use language effectively to perform and show mastery of typical communicative functions in specific contexts and not in the first instance on using linguistically accurate language. Since this is the approach that has been adopted for the design of language learning in South African schools, more attention will be devoted to the articulation of CLT in the prescribed curriculum in Chapter 4 which follows. Characteristic features on the language testing side are the use of texts and communicative task types extracted from real life contexts, i.e. "target language use domain" (Green 2014: 202). This should indicate a skills-neutral approach, since in authentic communication skills are used simultaneously. However, it seems that developers of language curricula and assessment artefacts cannot resist retaining the categories, entrenched by audio-lingualism, of teaching and testing reading, writing, speaking and listening separately, as any survey of authoritative textbooks in the field will reveal. Somewhat ironically the "integrated" forms of assessment being proposed also require combinations of open-ended and closed-ended items, with the latter assisting to provide reliable measurement and the former content validity (Weir 2005). In a bid for greater authenticity and construct validity, there seems to be a new emphasis on direct assessment of ability and the assessment of "productive skills" (Shin 2012: 239) that involve writing and speaking. Nonetheless, the psychometric approach introduced in first generation

applied linguistics continues to be useful, and multiple-choice item types are still used extensively in reputable tests of language proficiency across the globe.

The postmodern period does not signal a complete break with earlier traditions, but a shift in emphasis. One particular distinction is that language is viewed as dynamic and complex, and language learning as non-linear. Referring to the work of Kumaravadivelu (2006), Weideman (2017: 140) states that one of the things that can be learnt from the historical account of applied linguistics is that "such postmethod conceptualizations of language teaching designs confirm that the modernist search for the best method has been called off". This explains the eclectic approaches in use, since no method of language teaching or testing exists that can be sanctioned by science. Weideman comments on Kumaravadivelu's design approach:

> To its credit, this kind of conceptualisation much more realistically anticipates improvement in language teaching designs in terms of incremental gains that are for the most part locally conceived as well as highly contextualised. (Weideman 2017: 140-141)

We also see concerns being raised about values and consequences of a social nature, along with epistemological debates on the socially embedded nature of language and knowledge. The extent to which language proficiency should continue to be conceptualised (and assessed) with its current individualistic focus is also coming under scrutiny (McNamara 2005: 776). How the social and interactional dimension of language use can be incorporated in formal testing contexts remains to be seen. This rethinking of the social dimension in language teaching and testing resonates with fifth generation applied linguistics research on the socially constructed nature of knowledge, which in turn was preceded by second language acquisition studies (fourth generation applied linguistics) that focused on the negotiation of meaning and construction of knowledge while interacting socially. In the postmodern tradition we also see criticism being levelled at standardised and institutionalised language testing that has undesirable social effects, especially when used to promote political agendas (Shohamy 2001a, 2006; Fulcher & Davidson 2007; Xi 2010; Young 2012).

Although the focus of the current study does not fall only on the social or political effects of the HL examinations, the justifiability of the HL papers in their current formats and the extent to which they may be considered fair and trustworthy measurement artefacts

are pertinently relevant points, as was argued in Chapter 1. As already mentioned, prior to the 1960s little attention was devoted to whether the forms of assessment employed in language testing were beneficial and credible. Spolsky (1995) and Weideman (2017) argue that in order to avoid some of the language testing pitfalls encountered in the past century, educational authorities (and their appointed teams of examiners) should rely on more than theories underpinning only language, education and psychology. Insights gained from fields such as sociology, economics and political science can be of assistance to understand the constraints imposed on measuring language ability:

> The language tester is expected to be responsible for and responsive to theories derived from two unrelated and fundamentally inharmonious fields, linguistics (which wants to describe language knowledge) and psychometrics (which hopes to measure it and other human attributes), and at the same time is directed and constrained by rival practical, institutional, economic, social, and even political demands. (Spolsky 1995: 4)

In terms of the above view, any language testing event will inevitably be a compromise based on competing criteria of a theoretical and practical nature. On the matter of which of the criteria should be given pre-eminence, Van Dyk (2010: 44) foregrounds the need for integrity and accountability, especially where tests are designed for high-stakes purposes such as access to further study, as is the case with the Grade 12 examination papers. Part of this endeavor concerns ensuring the absence of bias so that no test questions favour any particular subgroup of the language testing population, and equitable treatment so that all test takers have a comparable, and hence fair, opportunity. To achieve this, it is clear that the conceptualisation of a theoretical framework for the purposes of designing the HL papers demands cognisance of past and current trends in language assessment and insights obtained from applied linguistics as the founding discipline.

The discussion in this section of traditions in applied linguistics and their relation to dominant phases in language teaching and testing shows that overlapping does occur and that there is an "unresolved (and fundamentally unresolvable) tension between competing sets of forces" (Spolsky 1995: 354). Spolsky views this tension as desirable: rather than simplistically advocating one approach over another, the strengths of each can be harnessed. However, practical and ideological forces ultimately tend to dictate what test design approach is followed, and what compromises have to be made. In all this, the social accountability of language testers comes to the fore and they must accept "full responsibility for the inevitable uncertainty of a powerful but flawed technology, and

make sure not just of reliability but also of focused and relevant validity" (Spolsky 1995: 358).

The conclusion to be drawn from the aforegoing discussion of traditions of applied linguistics and historical trends in language testing is that a multiplicity of perspectives and methods of assessment should be considered and, if useful, accommodated. Differentiation is important, as language test designs need to use a well-differentiated set of tasks to assess language ability adequately, a principle of test design that will again be referred to below. Each of the different traditions referred to developed from inadequacies identified in the preceding, whilst still retaining certain aspects of what was conventional before. Weideman refers to this as part of the innovation and reciprocity that characterises applied linguistics:

> We can continue to be surprised by innovation in the designed solutions that our profession provides, but we should also work on our understanding of what constitutes a responsible design framework. That foundation enables us to evaluate both the fleeting and the enduring in the new. (Weideman 2014: 1)

The above developments in the field of applied linguistics and language testing are relevant and their utility should be reflected in the format and task types of the national language examinations. Of particular relevance is the need to ensure both validity and reliability of assessment, as evidenced in the modernist and postmodernist traditions of language testing. The detailed content analysis of the English HL papers in Chapter 5, below, and structural comparison of a selection of English, Afrikaans and Sotho papers in the first part of Chapter 6, should shed some light on the extent to which current thinking is being reflected.

The next section discusses the core qualities that contribute towards the effective and responsible assessment of language ability within a postmodernist paradigm that values fairness and accountability in language testing.

3.4 Essential qualities of applied linguistic artefacts designed for the responsible assessment of language ability

The discussion on traditions in applied linguistics and language teaching has shown that the way language ability is conceptualised and assessed cannot be divorced from what a society values. Language testing (and other forms of tests) constitutes a means by which

values of a social, economic and political nature are realised. Where the purpose of examining language ability is aimed at providing "equality of opportunity", Fulcher (2010: 4) makes it clear that "meritocratic practices" should be entrenched in the assessment process:

> This means that we cannot separate the actual practice of writing tests and assessments – the nuts and bolts of test design and creation – from our values. For teachers and other practitioners, this is liberating. It means that our philosophy and understanding of what is valuable and meaningful in society and education are highly relevant to the tests that we use. (Fulcher 2010: 5)

In light of the fact that the qualities of language tests reflect the values of a particular community, it can be postulated that examination papers which do not comply with those qualities which have been shown to be beneficial and socially just, reveal a negligence on the part of those responsible for their design and use. Although this negligence may not be conscious or deliberate, its impact on all stakeholders can be negative. This investigation will show how it is thus of paramount importance that Umalusi treat the design of the HL papers as a matter of priority and initiate steps to ensure that the papers reflect the necessary principles for responsible design.

Literature shows that discussions on the essential qualities of professionally designed language tests revolve primarily around the notions of validity and reliability (Alderson, Clapham & Wall 1995; Spolsky 1995; Cumming & Berwick 1996; Hughes 2003; Weir 2005; Weideman 2009a; Fulcher 2010; Van Dyk 2010), as well as issues of fairness, accountability and practicality (Bachman & Palmer 1996, 2010; Kunnan 2000; Fulcher & Davidson 2012; Rambiritch 2012a; Green 2014; Fulcher 2015). The term "validation" (or responsible design) then incorporates all of the above-mentioned aspects and is used with reference to the process to collect evidence that the design and administration of a test has been useful to produce scores from which credible inferences about specific aspects of language ability can be made (see Weir 2005: 1, Fulcher 2010: 324; Saville 2012: 404; Green 2014: 76). In view of the fact that the current study focuses mainly on the design format and content of the HL papers, and not specifically on the administration of the examination or its consequences, the elements of validity, reliability and practicality will be most relevant for the purposes of the discussion. The next section discusses how these three internationally recognised principles in language testing all contribute significant but different forms of evidence towards the validation argument for the kind of language assessment that is the focus of this study.

3.4.1 Understanding validity

Much attention in the last few decades has been devoted to alternative conceptualisations of validity. Early views of validity were limited to "the extent to which an examination actually measures what it purports to measure" (College Entrance Examination Board 1937: 31 in Spolsky 1995: 83; also see Lado 1961: 30). The fact that this is an oversimplification of a phenomenon that is difficult to determine is evident in Lado's (1961: 20) explanation that time constraints necessitate testing in a brief hour (or two) what has been learned over a long stretch of time, and using only a sample of work to do so. Lado's strategy to ensure the validity of test items is, however, problematic. He advised eliminating aspects already known, and focusing on problematic areas: "We will attempt to test the learning problems, on the grounds that knowing the problems is knowing the language. We say specifically that testing the problems is testing the language" (Lado 1961: 25). By language, Lado meant aspects such as sound and intonation, stress, morphemes and word arrangements, in tandem with the structuralist approach to language teaching prevalent in the early tradition of applied linguistics, and the technique of contrastive analysis it spawned.

A further problem with Lado's approach is that he associated validity with subjectivity. In his view, tests that required subjective scoring allowed the "use of techniques that are natural and seem outwardly very valid … Often we have to choose between more apparent validity but less objectivity and more objectivity but less apparent validity" (Lado 1961: 29). According to Lado the dilemma can be resolved in part by accurately describing the linguistic problems to be tested through linguistic analysis:

> Knowing what an item is testing we are left free to choose objective techniques even if outwardly they seem less valid, for if they test the language problems in essentially valid linguistic situations they are valid items. (Lado 1961: 29)

In Lado's conceptualisation, validity and objectivity are at times opposite and competing principles. Few contemporary language testers are likely to agree with his view on validity, since today objectivity is considered to be an important contributor towards reliability, which in itself is a distinct but essential condition for validity. Also contestable would be his view of what should constitute the test content. Focusing on what learners are not likely to know ("learning problems") would be considered biased and unfair in a

postmodern paradigm. What is more, what counts as "learning problems" is defined strictly within the conceptualisation of language within the contestable structuralist view of language underlying contrastive analysis.

Validity theory, however, has moved on since Lado. The fact that validating a test is an extensive process is made clear by Angoff (1988) who mentions that evidence of validity is to be found in all data generated by a test. Many different kinds of validity are prevalent in psychometric literature dating from the 1930s. Those traditionally cited in language testing literature include the following (based on Cumming & Berwick 1996: 1-12; Hughes 2003: 26-35; Weir 2005: 11-15):

- Criterion-related validity: a combination of concurrent and predictive validity and an aspect that pertains to the correspondence of one set of test results with the results of another test of ability that serves as a criterion
- Concurrent validity: the ability of the results of one test to correlate with those of another criterion believed to indicate the same ability as that which has been tested
- Predictive validity: the ability of the results of one test to predict performance in other situations and contexts
- Construct validity: the theoretical trait or construct of a cognitive and linguistic nature presumably measured, and the alignment thereof with theories on language processing and communicative competence
- Content validity: the adequate representation of authentic language-related tasks and content in a test

Since Messick (1988) drew attention to the consequences of testing, the usefulness and relevance of test scores have been foregrounded and the emphasis placed on construct validity more than the other types. In the process a broader interpretation has been accorded to construct validity, which has come to be seen by many language testers as a uniform principle that "subsumes various other aspects of validation" (Cumming & Berwick 1996: 5). In this popular understanding, construct validity "integrates considerations of content, criteria and consequences into a comprehensive framework for empirically testing rational hypotheses about score meaning and utility" (Messick 1995: 742). According to this view, rather than considering validity to be an inherent property of a language test or examination, validity is viewed as "a function of the way in which the results can be meaningfully interpreted" (Read 2010: 288; see also Angoff 1988: 24; Kane 2004: 136; Chapelle 2012: 21; Van der Walt 2012: 145; Green 2014: 75). Further to this, validity should be considered as a "matter of degree" (Green 2014: 76), since inferences cannot be declared to be fully valid or invalid.

Some aspects of the Messick view are contestable, since it may be argued that validation commences *before* scores are generated and results are subjected to interpretation. After all, there are different phases in the measurement process and some properties of a test can be deemed to be theoretically valid prior to the administration of the test, while others may be found to have empirical validity after the analysis and interpretation of scores. For example, the statistical analysis of test scores may show scoring to be reliable, strengthening at least part of the argument for the validity of the usefulness of the test. However, should scoring be found to be unreliable, this would not necessarily affect the theoretical validity of the construct. Rather, the method of scoring used may need to be amended, or certain test items may need to be refined so as to improve the consistency of measurement. Weideman (2012: 4) expresses concern that making validity "dependent on interpretation runs the risk of downplaying the quality of the measurement" and that no amount of interpretation of results is helpful where the design of the test is in the first place inadequate.

Further to the above, if we consider the Messick view more closely, then another problem emerges. Messick (1989) states that validity is:

> An integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment. (Messick 1989: 13)

There can be little disagreement that validation is a complex process that continues beyond test administration and that requires several "evaluative" judgements to be made. However, at what point can it then be decided that the evidence obtained in the different stages of test development and administration is now adequate? To complicate matters further, in keeping with postmodernist thinking, the results of one validation process may not be suitable for other assessment contexts. Weideman (2009a: 236) responds to this part of the Messick debate by stating that evidence is *necessary*, but it may not always be possible to find evidence to be *sufficient*. If we take into account the fact that Messick's formulation of validity pertains to educational measurement in general, this may explain why it cannot be operationalised in the same way for the measurement of language ability which is not a stable trait of a quantitative nature. Could it not be argued that if the test construct is appropriate for the purposes of the assessment and has been articulated adequately through sufficient corresponding tasks and items, there can be little reason to doubt the adequacy and appropriateness of the evidence of ability inferred on the basis of

test scores? Only an illogical interpretation of scores could undermine the latter. Moreover, the impracticality of considering test consequences to form part of validation is obvious too. We simply cannot foresee all the effects of tests.

From the above it is obvious that a logical place to start would be to reach clarity on what is to be assessed, i.e. the construct, before proceeding with the design of the measurement artefact so as to ensure that the objectives of the assessment are executed in a responsible way. As Weir (2005: 18) points out, statistical data are unable to generate concepts, and test designers carry an obligation both to define what they want to measure and to investigate the adequacy of their measurement. The test (or examination in the current study) serves as an indirect way to describe to what extent a test taker or examinee displays the ability that has been conceptualised theoretically (ALTE 2005: 21). If we view validation as a process, it commences with the purpose of the assessment and articulation of the construct, and in this sense construct validity provides a central point of departure. Whereas the term construct refers to the definition of an ability that is to be tested, construct validity pertains to the degree to which a given test score can be interpreted as a valid indication of ability with reference to the definition of that ability. Van Dyk and Weideman (2004b: 17) refer to this as the alignment of the definition of the construct with what the testing instrument actually measures, i.e. the extent to which the defined concept is reflected in the test. However, by considering the social consequences and potential misuse of a test to constitute part of construct validity, as many contemporary language testers do, the theoretical definition of a construct is exceeded. The same applies to making validity dependent on interpretation of results, as in Messick's formulation quoted above, and its subsequent reformulations by others (see Weideman 2012).

Construct validity provides the necessary justification for the interpretation and generalisation of test scores. Since a generic and differentiated English language proficiency is the construct under consideration in the current study (as will become apparent in Chapter 4), this needs to be assessed with an enriched, open view of language rather than in terms of a mere four skills-based (listening, speaking, reading, writing) restrictive approach (Van Dyk & Weideman 2004a).

Patterson and Weideman (2013a: 108) provide further clarification on the term construct by explaining that it refers to a "theoretically defensible definition of what it is that should be measured". Much more is thus involved than merely identifying the underlying ability. There has to be a theoretical rationale behind it. Because of the importance of achieving clarity on the construct of a language test or examination, the next chapter is devoted to articulating the construct on the basis of the prescribed school curriculum. As will be seen later in the discussion, any articulation of the construct is dependent on the typical features of the discourse involved. Moreover, examinees will also differ in their ability to "handle each different type of discourse" signalling the differential ability (Patterson & Weideman 2013a: 108) that CAPS seeks to place at the basis of HL instruction, as the next chapter will indicate.

Messick's unified construct validity framework has also come under criticism for being complex and unmanageable to language teachers without formal training in assessment (Rambiritch 2012b; Van der Walt 2012). Although influential in psychometrics, it has not proven to be helpful in contexts where clear guidelines for the design of tests are needed. Rambiritch raises the additional concern that the Messick view relies heavily on empirical data and statistical procedures which may not be adequate to account for all aspects involved in measuring language ability. A third problem that she mentions is the inability of Messick's validity matrix to resolve social issues relating to testing.

Rather than advocating a unitary or superordinate notion of construct validity, language testers such as Bachman and Palmer (1996, 2010), Weir (2005), Weideman (2009a), Van Dyk (2010), Rambiritch (2012b) and Green (2014) place the emphasis on a number of essential test qualities that should feature in the test validation process, of which validity is one. Weir (2005: 13) furthermore states categorically that no "single validity can be considered superior to another. Deficit in any one raises questions as to the well-foundedness of any interpretation of test scores". This resonates with the view of Bachman and Palmer (1996: 9) whose validation framework includes six variables for the usefulness of a test, of which construct validity is one; inadequacies in any of these would undermine the usefulness of the test in their opinion.

3.4.2 Beyond the orthodox notion of validity

Drawing on his paradigm of applied linguistics as a discipline of design, Weideman (2009a) provides a framework for language testing based on two main tenets of an analytical and theoretical nature which he refers to as constitutive and regulative conditions. Rather than attempting to subsume any of the necessary qualities under a single unitary notion such as validity, he discusses each as being of relevance and interrelated, cautioning that the "requirement of conceptual acuity for the sake of an improved designed instrument is not served if concepts are conflated" (Weideman 2009a: 241). In his opinion conflating everything under a unitary concept of construct validity is not only unnecessary, but undesirable. Weideman (2017) therefore prefers the idea of "responsible design" to the concept of validity in this respect.

Weideman (2009a) agrees with Kane and others that scores are meaningless objects without human interpretation, but cautions that objective measurements are used to make subjective interpretations. No interpretation is possible without a score, which is the objective result of the measurement, and the reverse is also true: no objective measurement has any sense apart from its legitimate interpretation. There is thus a distinction to be made between the "subjective process of validation and the objective validity of a test" (Weideman 2009a: 243). As such, it would be preferable to speak in process terms of a validated test rather than a valid test. Validation should also be described as a matter of degree, since interpretations are of a subjective nature and it would be hard to argue a case for every aspect to be validated in full. The conditions that in this broadened perspective should form part of a framework for responsible design are presented graphically in Figure 3.1.

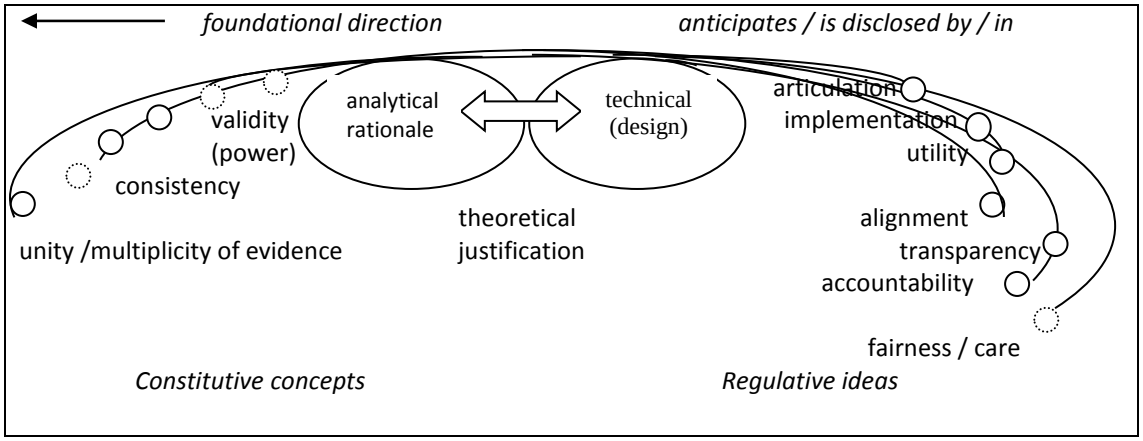


Figure 3.1: Constitutive and regulative conditions for the validation of language tests (Weideman 2009a: 248)

In terms of the above representation, the theoretical justification for the design of a language test is to be found in the reciprocal relationship between the analytical and technical modes. In language testing the technical (design) mode leads and qualifies the design of a solution to a language related problem, while the analytical dimension provides the foundational basis for the intervention. This view is shared by Van Dyk (2010) who states that it is through the process to validate a test that a sound theoretical foundation for the designed artefact is provided with a view to ensuring the fairest outcome for the test taker. The respective normative moments reflected in Figure 3.1 will be discussed later in this chapter.

In his framework for validating tests, Weir prefers to speak of an evidence-generating process that involves two dominant phases. The first of these, which he refers to as the generation of *a priori* validity evidence, takes place during the planning and design of the language examination or test. The second phase, during which scoring and interpretation of results take place, is used to generate *a posteriori* validity evidence (Weir 2005: 17). However, when viewed as a process, in addition to the two dominant phases identified by Weir, validation should include a third intermediate phase which may be referred to as the generation of *per administratio* validity evidence during the actual administration of the test or examination. It may be argued that without uniform conditions for sitting a test, the validity and comparability of the results could be undermined. This aspect is nonetheless often beyond the control of test designers, which may explain why little mention is given to the conduciveness of the institutional and organisational environment for assessment purposes. In the case of the high-stakes assessments that are the focus of

this study, however, the standard and uniform administration of the examination is a crucial factor in its fairness.

Weir further distinguishes two forms of validity that he considers intrinsic to the notion of construct validity as part of generating *a priori* evidence. He refers to these as theory-based validity and context validity (Weir 2005: 17-19). Theory-based validity relates to the necessity of showing that a test or examination "correlates highly with indices of behaviour that one might theoretically expect it to correlate with" (Weir 2005: 18). Stated differently, the underlying language and cognitive processing that takes place when performing language related tasks in real-life contexts needs to be replicated in the operationalisation of the sub-skills to be measured. This is the principle referred to by Bachman and Palmer (1996) in their framework as interactiveness, which they explain is

> … a function of the extent and type of involvement of the test taker's language ability (language knowledge plus metacognitive strategies), topical knowledge, and affective schemata in accomplishing a test task. (Bachman & Palmer 1996: 39)

These three characteristics of the test takers affect the way they interact with the test tasks. Tasks that require test takers to relate topical content to their own topical knowledge can be expected to be more interactive. According to Bachman and Palmer (1996: 26) the element of interactiveness is an essential quality of any language test task, because it provides the necessary link with construct validity.

The second kind of validity intrinsic to construct validity according to Weir is generally referred to as content validity, but he prefers to speak of context validity so as to reflect a socio-cognitive approach to language testing. He describes context validity as "the extent to which the choice of tasks in a test is representative of the larger universe of tasks of which the test is assumed to be a sample" (2003: 19) with due consideration to the "linguistic and interlocutor demands made by the task(s) as well as the conditions under which the task is performed", i.e. the notion of authenticity discussed in Bachman and Palmer's (1996) framework and the alignment of test tasks with target language use (TLU) activities in real social settings.[12] Although Weir considers construct validity to include the operationalisation of the construct in the selection of language tasks, those

[12] Notwithstanding the necessity of alignment of test tasks with TLU domains, attempting to replicate natural language use in an artificial test setting will render all forms of language assessment inauthentic to some extent, a reality that should be recognised by test developers and users.

advocating for a broader understanding of the Messick view of validity would consider the selection and relevance of task types as giving effect to the construct and not affecting its validity. In other words, the selection of unsuitable task types would not alter the objective of the assessment, but would render the methods used to measure the articulated ability ineffective. Construct validity focuses on the "what" that needs to be measured, and content/context validity on the "how". This explains why in the framework of Weideman (2017) construct validity features under the analytical analogy, and the ability to differentiate is understood to be a biotic analogy. This is in keeping with dynamic systems theory (cf. Weideman 2017: 166ff) according to which language and the use thereof are dynamic and unpredictable depending on the interaction of many components. (This distinction is reflected in the comprehensive framework reflecting the normative moments in language test design presented in Table 3.5 towards the end of the chapter.)

As part of establishing *a priori* validation evidence, theoretical and context validity need to be delineated from the onset of the process to design the language test or examination, which corresponds with the view of Weideman, Van Dyk and others that it is through the articulation of the construct that the theoretical justification for the design is found. We can see that the validity of using a test for a specific purpose can only be inferred when the abilities to be assessed can be generalised to non-testing contexts and are founded on accepted theories of language, cognition and communicative competence. In this respect, it is essential that aspects not related to the abilities or knowledge to be measured do not form part of the test construct (i.e. construct irrelevant features are excluded), and that test tasks adequately represent the intended construct (i.e. underrepresentation of the construct is prevented). Messick (1989: 34) identifies the latter as two aspects that threaten validity. To counter this, enough task types need to be selected and these need to be suitable for measuring the specified abilities or knowledge.

This section has dealt with a broadened perspective on validity and validation, and how that expands the orthodox perspective on this, as well as how that may be applied in assessment arrangements and designs. The following section discusses the close link between validity and reliability and how inconsistency of measurement can undermine the validity argument.

3.4.3 The condition of reliability

Reliability in language testing refers to the degree to which scores obtained in a test display consistency of measurement (Bachman & Palmer 1996: 19; Green 2014: 63), and the "proportion of variation in scores caused by the ability measured, and not by other factors" (Jones 2012: 352). Test results are deemed to be reliable if they remain consistent from one set of tests and tasks to another. Reliability is thus a function of score consistency between different administrations of tests and tasks. In other words, a test taker should obtain a similar score if the same test is administered to the same group of test takers on two separate occasions and settings. Reliability is essential – a necessary condition for assessment design – if a test score is to provide credible information about a test taker's language ability and if the results are to be generalised to non-testing domains.

Reliability in language testing may be more difficult to determine than in the case of educational measurement in general. Language proficiency is a construct that cannot be articulated or measured with absolute precision, because it is not a stable trait of a quantitative nature, but a phenomenon "situated in the variability of social use and interaction" (Jones 2012: 351). In addition hereto, it is understood in different ways, making it difficult to compare examination results across different administrations of tests and across different language groups. Further to this, there are a number of factors that can impinge on the outcomes of the assessment which serve as potential sources of measurement error. Jones (2012: 352) summarises these as factors pertaining to the individual test takers (health, motivation, etc.), testing situation (environment), persons responsible for marking (severity, training), and instrumental factors directly related to the test itself (content, technical issues, etc.). In the current study, the first two factors are beyond the control of the test designers and Umalusi, which means that the emphasis needs to fall on the system of marking used and the language papers themselves as the assessment instruments.

With regard to the scoring of written answers, in light of the fact that marking of the Grade 12 HL papers is carried out manually by thousands of evaluators at different centres across the country, the potential for error of measurement is obvious. Irrespective of whether scoring takes place manually or with the aid of technology, every attempt must

be made by those responsible for designing the HL papers to ensure that examination tasks can be scored as reliably as possible. Where samples of writing are scored by different persons, as in the case of Paper 3, the consistency of the marking needs to be estimated. Weir (2005: 34) points out that markers need to have high reliability in terms of the consistency with which they mark ("intra-rater reliability"), as well as in terms of the severity of their marking as compared to that of the other markers ("inter-rater reliability"). A correlation of at least 0.9 is desirable if the scoring is to be considered valid. Umalusi thus needs to ensure that use is made of processes and techniques that can increase the consistency of the marking of HL papers. As Green (2014: 64) states, without "reasonable estimates of the amount of error", at best we can guess what the language ability of test takers is since we cannot determine how accurate the scores are likely to be. Detailed rubrics, rating scales and a system of rater training and moderation are recommended to increase the reliability of scoring. Although Umalusi does employ these, the reliability of the current system of scoring has not been verified empirically.[13] What is more, even in a highly moderated system with highly qualified markers at one South African university where language tests have been designed since 1989, there is seldom an inter-rater correlation of higher than 0.6 (Weideman, personal communication). It is therefore highly unlikely that the current Grade 12 examination arrangement, undertaken by less qualified markers, will be anywhere near that, which may explain Umalusi's reluctance. However, the technology exists, for example in the employment of a programme such as Facets, not only to determine and compare rater severity or leniency, but also to compensate for such deviation by recalculating marks using Rasch analysis.

In respect of the instrumental factors pertaining to the HL papers themselves, reliability can be estimated through statistical procedures that quantify the likelihood that a test score will diverge from the true score (see Hughes 2003: 36 ff.; Jones 2012: 354 ff.). This may be undertaken by applying either classical test theory (CTT) or item response theory (IRT) that usually includes Rasch analysis (McNamara 1996). While the former method generates a single estimate for all candidates, the latter provides an estimate for each individual's true score. Furthermore, inferential statistical analyses make it possible to detect inconsistencies in test items and to identify items that perform poorly and do not

[13] Attempts to obtain examples of examined Grade 12 scripts from Umalusi have been unsuccessful so far. This should be pursued in a further study. The capability of appointed markers has come under severe criticism (see Loggenberg 2013).

discriminate well between candidates of differing ability. Not all task types are equally productive and the items of each type can be considered to fall into one of four categories:

> *acceptable* (a high degree of alignment with the test construct, but apparently not productive), *unacceptable* (low productivity coupled with small degree of alignment with blueprint), *desirable* (high alignment with construct, as well as productive), or *not ideal* (potentially productive, but not quite aligned with framework. (Van Dyk & Weideman 2004b: 17-18)

The ideal is to include only, or at least as many highly productive test items as possible, such as in the case of professionally designed standardised tests such as the NBTs and Test of Academic Literacy Levels (TALL) (ICELDA 2015). Where there is a need to ensure consistency of content and test administration, standardised tests are usually employed (Green 2014: 28). In essence, standardised testing refers to the administration of the same test under the same conditions to different groups of candidates, as well as clearly defined and fixed methods of scoring (Fulcher 2010: 5). Normally those items to be included in a standardised test would undergo a process of refinement based on the inferential statistics generated in pilot tests, so as to achieve optimal productivity before being item banked for future use. Standardised testing applies to the current HL papers in so far as all students write the same (nationally set) examination per language group. However, no use is made of piloting of items and item banking. Although Umalusi could benefit from incorporating technology to a greater extent as part of the design of the HL papers and administration of the HL examinations, and although some of the reports they have commissioned to address their unease with the current set of HL papers have proposed item banking after pre-testing, no action to put this in place has yet happened.

Brown (2012b: 326) points out that currently CTT is still the main psychometric theory applied to ensure reliability in language testing, and that it "serves as the basis for understanding all aspects of language testing". Although there is interest in IRT, so far it has not been used extensively owing to practical disadvantages and accessibility issues (Ockey 2012: 347). In CTT, use is made mainly of descriptive statistics, determining the standard error of measurement, reliability estimates and item analysis. These aspects will be discussed briefly next. In IRT, though it is suitable only for the analysis of larger groups, there is, however, the gain of going beyond descriptive and inferential statistics to probabilistic models that allow a further measure and level of generalisation.

3.4.3.1 Descriptive statistics

Descriptive statistics conventionally provide the basis for interpreting other statistics needed to generate a validity argument (Brown 2012b: 332; Bachman 2004: 33). These include the production of frequency tables and histograms, measures of dispersion and measures of central tendency (Hughes 2003: 218-222).

Histograms and frequency tables are useful to see the spread of scores obtained and the range, i.e. the "difference between the highest and lowest scores" (Bachman 2004: 63). The range is useful for noticing to what extent the levels of ability vary amongst the test population (Fulcher 2010: 37), an aspect which can also be used to make comparisons across HL examination populations. The HL papers are norm-referenced, which means that scores of individual examinees are compared relative to those of other examinees. If the distribution is normal, a bell-shaped curve is produced. Most scores should fall close to the mean of the curve (68%), with around 34% just under the mean and around 34% just over the mean:

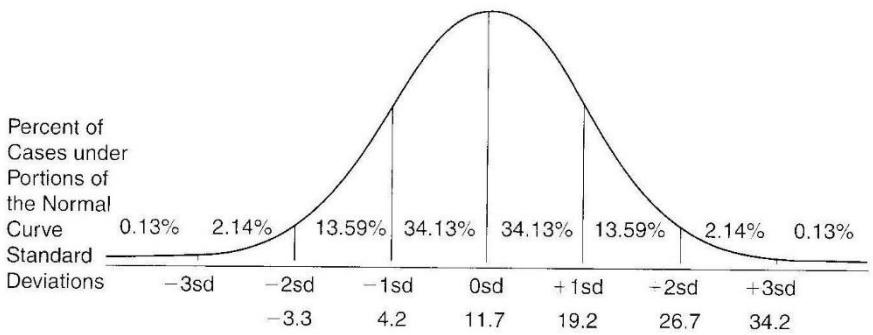


Figure 3.2: Normal distribution curve with raw scores (Fulcher 2010: 40)

The objective of examining the normal distribution is summarised by Fulcher as follows:

> The curve of normal distribution tells us what the probability is that a test taker could have got the score they have, given the place of the score in a particular distribution. And this is why we can say that a score is 'exceptional' or 'in the top 10 per cent', or 'just a little better than average'. (Fulcher 2010: 37)

To function effectively for the purposes of score interpretations, the HL papers need to generate scores with a wide distribution (Bachman 2004: 30). Currently, the final mark per examination candidate is obtained by combining the results of HL papers 1-3 and the

score per examinee obtained for school-based continuous assessment. The final composite score is reported on the basis of a rating code and is used for making decisions as to which examinees pass or fail, and for university or college admission purposes.

Table 3.3: Codes and percentages for recording and reporting in Grades 10-12 (Department of Basic Education 2012e: 21)

RATING CODE	ACHIEVEMENT DESCRIPTION	MARKS
7	Outstanding Achievement	80 – 100
6	Meritorious Achievement	70 – 79
5	Substantial Achievement	60 – 69
4	Adequate Achievement	50 – 59
3	Moderate Achievement	40 – 49
2	Elementary Achievement	30 – 39
1	Not Achieved	0 – 29

The mean (commonly referred to as average score) obtained per administration of each HL paper is used to indicate the central tendency (Bachman 2004: 63) and is calculated by adding all the examination scores together and dividing by the total number of examinees. It was on the basis of an interpretation of the mean obtained per HL subject over a four-year period (Table 2.3) that Umalusi realised that the HL papers may not be comparable in terms of standard or construct.

The mean is also used to calculate the standard deviation, another measure of dispersion which indicates "how much, on average, test scores tend to vary, or deviate, from the mean" (Bachman 2004: 66). The standard deviation is calculated through first calculating a deviation score (x) for each examinee by subtracting the mean ($\overline{X}$) from each raw score (X). Each deviation squared is multiplied by its frequency of occurrence. Then the sum (Σ) of deviations squared is obtained. The standard deviation (SD) is thus calculated by obtaining the "square root of the sum of the squared deviation scores, divided by N-1" (Fulcher 2010: 40; N = number of scores):

$$\mathrm{SD} = \sqrt{\frac{\Sigma(\mathrm{X} - \overline{\mathrm{X}})^2}{\mathrm{N} - 1}}$$

The standard deviation is important because it is used to determine the standard error of measurement (SEM), another indicator of the reliability of individual test scores.

3.4.3.2 Standard error of measurement (SEM)

No language test can be fully reliable because language ability cannot be observed in "consistent, predictable ways" (Fulcher 2010: 4). As Bachman (2004) explains:

> … scores are reliable to the extent that they are free from measurement error. Or, to put it another way, test scores that reflect mostly the ability we want to measure, rather than measurement error, will be relatively reliable. (Bachman 2004: 153)

Any test or examination score (X) will consist of the "true" score (T) representing the ability of the examinee, as well as the error (E) deriving from various sources (Fulcher 2010: 47):

$$X = T + E$$

The SEM is needed in conjunction with other reliability estimates to "provide information for interpreting the effects of measurement error" on the scores obtained by examinees (Bachman 2004: 171). The formula used to calculate the SEM requires the standard deviation and the reliability coefficient (R), which indicates the amount of error that there might be (Fulcher 2010: 54):

$$\mathrm{SEM} = \mathrm{SD}\sqrt{1 - \mathrm{R}}$$

The calculation of the reliability coefficient (R) will be discussed in the next subsection. Through calculating the SEM, a confidence interval can be generated in respect of any observed score to indicate how much this score may fall below or above the observed score. In language tests the confidence interval is normally 95%, i.e. "we can be 95 percent confident that the true score will fall within a certain range above and below the observed score" (Fulcher 2010: 55). Moreover, the "95 per cent confidence interval falls at 1.96 standard deviations, and the 99 per cent confidence interval at 2.58 standard deviations" (Fulcher 2010: 55). Use is made of statistically generated z-scores (raw scores expressed in standard deviations) to indicate the proportion of scores that fall between a z-score of 1.96 and the mean (see Fulcher 2010: 301). For example, in a language test with a SD of 7.5, the SEM will be 1.7. If the SEM is multiplied by the 95 percent level of certainty, the following indicator is obtained (Fulcher 2010: 56):

95% confidence interval = 1.7*1.96 = 3.33

This means that for any observed score the test users can be 95 per cent confident that the score will lie "within a range of + or – 3" (Fulcher 2010: 56). In other words, for a test taker who obtained a test score of 14, the true score is likely to be between 11 and 17. When making decisions on pass marks or cut scores, the SEM is clearly an aspect that needs to be taken into consideration.

3.4.3.3 Reliability estimates

Initially, reliability was estimated through comparing the results of tests that were repeated or those of similar tests written by the same test takers (test-retest and parallel form methods). However, owing to the practicality with which they can be used, internal consistency estimates have largely replaced these earlier methods, and today only data from a single test administration need be analysed (Jones 2012: 354).

Reliability in CTT is defined on the basis of true score variance and error variance (Brown 2012b: 325). When scores on a test vary, some of the variance can be attributed to the construct of the test. As pointed out earlier, "true scores are hypothetical representations of the scores" in the absence of errors in measurement. The variation that occurs among the hypothetical "true" scores of examinees is referred to as "*true score variance*" (Brown 2012b: 323). Any variation in the scores that is not systematic and that is the result of random errors is known as "*error variance*" (Brown 2012b: 323). Reliability refers to the true score variance divided by the "total variance" on a test, i.e. the true score variance added to the error variance (Brown 2012b: 326):

$$\text{Reliability} = \frac{\text{Var}_{\text{true score}}}{\text{Var}_{\text{true score}} + \text{Var}_{\text{error}}} = \frac{\text{Var}_{\text{true score}}}{\text{Var}_{\text{total}}}$$

There are a number of statistical techniques to measure reliability, such as split-half reliability estimates, the Kuder-Richardson (K-R) 20 formula, and Cronbach *alpha* (also referred to as Cronbach's *alpha*). Programmes such as Iteman 4.3 (Guyer & Thompson 2011) can be used with considerable ease to provide such estimates.

Split-half reliability separates the items in a test to create two separate forms of the test. A correlation coefficient is obtained (typically between even-numbered and odd-numbered items). This estimates the reliability of two halves of the test separately, which will be lower than the reliability of the full test. The half-test correlation is adjusted on the basis of the "Spearman-Brown prophecy formula for a test of double the half-test length" (Brown 2012b: 328), because of the direct relation between reliability and test length:[14] Fulcher (2010: 51) provides the following formula in which r_{hh} represents the correlation between the two test halves:

$$R = \frac{2r_{hh}}{1 + r_{hh}}$$

The K-R20 formula is commonly used because it is easy to calculate with computer technology. Brown (2012b: 329) explains that this formula requires the number of test items (k), standard deviation (SD), proportion of examinees who provided the correct answer per item (i.e. item facility or p), and the proportion of examinees who provided the incorrect answer (q) (Brown 2012b: 329):

$$K - R20 = \frac{k}{k-1}\left(1 - \frac{\Sigma pq}{SD^2}\right)$$

To calculate Σpq, it is necessary to calculate the product of p and q for each item and then determine the sum of all pq values for all items, which can be tedious (Brown 2012b: 329). One of the problems with the K-R20 formula is that it requires items to be scored dichotomously (i.e. an answer is either right or wrong). A formula that provides more flexibility is that of Cronbach *alpha* (α) which can also accommodate weighted items (i.e. items that count more marks than others). It should be noted that where items are scored dichotomously, α is calculated using the same K-R20 formula provided above. However, where rating is used to score a test item, the variation between scores needs to be quantified since some raters may be inclined to score more leniently than other raters.

[14] Longer tests have been shown to have higher reliability (see Geldenhuys 2007; Green 2014).

Inter-rater reliability between two different raters can be calculated using the following formula (Fulcher 2010: 53):

$$\alpha = \frac{k}{k-1}\left(1 - \frac{S^2{}_{r1} + S^2{}_{r2}}{S^2{}_{r1+r2}}\right)$$

In the above formula, "k" represents the number of raters, while "S^2" shows the variance of the raters' scores. The different raters are indicated by means of "r1" and "r2". A further method that can be used to identify inconsistent marking is Multifaceted Rasch (MFR) analysis (see Bachman 2004: 146 ff.; Weir 2005: 199 ff.), which forms part of Item Response Theory (IRT) – an "approach used to estimate how much of a latent trait an individual possesses" on the basis of mathematical modelling (Ockey 2012: 336).

In summary, Cronbach *alpha* shows the "degree to which the observed scores represent the 'true' scores", without measurement error (Van der Slik & Weideman 2005: 26). Bachman and Palmer (1996: 135) consider the purpose for which the test is intended as probably the most important aspect when determining a minimum acceptable level of reliability. For a high-stakes test the minimum acceptable level of reliability should be set as high as possible. A Cronbach *alpha* of around 0.7 is considered to be suitable for classroom testing and research purposes (Hogan 2007: 149-150; also see Van Dyk 2010: 154). Tests such as TALL require a Cronbach *alpha* of at least 0.8, which means that 80% of the total variance can be considered true score variance, and 20% is error variance. Even if a standardised test is used, Brown (2012b: 330) cautions that reliability estimates are always linked to a particular group of examinees. Reliability estimates should thus be interpreted with reference to a particular group and specific test items. Nonetheless, the usefulness of obtaining reliability estimates for the respective HL papers for the purposes of comparing performance across languages is obvious, and the fact that it is not yet being done constitutes a serious impediment to judgements of their fairness.

It should further be kept in mind that reliability is harder to achieve when the construct is complex and covers a range of language ability components and topical knowledge, such as in the case of the HL papers. Item analysis, which I shall discuss next, can provide useful information to supplement the reliability estimates.

3.4.3.4 Item analysis

Internal consistency or correlation coefficients are determined statistically through programmes such as Iteman 4.3 (Guyer & Thompson 2011) and Tiaplus (CITO 2005) and include the calculation of facility and discrimination values, as well as distractor efficiency estimates where multiple choice items are used (Brown 2012b: 326). Often the reason for the weakness of a test item can be found in one of these. Facility values reflect the percentage of correct answers for the whole of the test population. A well designed test should produce a normal distribution of scores, indicating that the number of items that are very difficult or easy is not disproportionate to the facility values of the remaining test items. Facility values of between 0.30 and 0.70 are recommended in CTT since they tend to discriminate well between candidates of differing ability (Brown 2012b: 327).

How well a test item can distinguish between strong and weak test takers is referred to as the discrimination index. Items with high discrimination indexes make the test more reliable (Le 2011: 22). Item discrimination provides an indication of the extent to which the performance of a test taker on a particular test item relates to performance on the whole test. A correlation coefficient is calculated to show the "strength and direction of the relationship" between two variables (Bachman 2004: 84). The Pearson point-biserial correlation (*r-pbis*) is normally used in language testing to measure the differentiating strength of a test item. The correlation can range from -1.0 to 1.0, but does not usually exceed 0.50 (Guyer & Thompson 2011: 30). If a negative point-biserial is obtained, this shows that the item is a poor one and that strong candidates are answering it incorrectly, while weaker test takers are providing correct answers. If the point-biserial is 0, this means that the item does not provide any differentiation between low and high scoring test takers and that this item either needs to be refined or rejected.

When multiple choice items are used in a language test, the effectiveness of the choices of answers provided can be determined through statistical analysis. For example, incorrect answers ("distractors") that are not being selected at all by test takers are deemed to be inefficient (Brown 2012b: 327) and can be substituted with more productive choices in future test versions.

Internal consistency estimates may, however, be less relevant when applying integrated forms of assessment that require multiple and heterogeneous abilities to perform test tasks. A scatterplot depicting a factor analysis is used to illustrate to what extent the test is one-dimensional or multi-dimensional, i.e. whether one or more abilities are being tested at the same time. This is depicted by the way in which the items of test sections cluster in more or less the same range. A rich construct such as high language ability is likely to indicate a certain measure of heterogeneity and that more than one ability is being assessed simultaneously. Nonetheless, high internal consistency can still be achieved where the test is multidimensional.[15]

3.4.3.5 The importance of reliability

The preceding discussion shows how reliability or consistency of measurement can contribute towards the construction of a validity argument. However, it should be emphasised that although reliability is a necessary quality of a language test, it is not an aspect on its own that can provide sufficient evidence of the validity of the measuring instrument. Weir (2005: 24) in fact somewhat ironically refers to scoring validity as a superordinate term for reliability. Once again, this kind of conflation of different aspects under an umbrella term only serves to obstruct conceptual clarity. It may be argued that there is a distinction to be made between reliability resulting from valid scoring procedures and norms, and reliability resulting from the productivity of test items that consistently distinguish between differing levels of ability across different test administrations. Scoring validity would thus not be an appropriate term for any superordinate category of reliability. Nevertheless, it should be recognised that the design of scoring rubrics may be considered an essential part of the endeavor for construct validity, since there needs to be alignment between the articulation of the test construct and the articulation of language ability in the rubrics that will be used to measure ability.

As has been observed above, to date the education authorities have placed little emphasis on the condition of reliability in language assessment. Van Dyk (2010: 119) points out the postmodern tendency to shift the focus away from objectivity or consistency of

[15] There are also acceptable parameters for the inter-correlations between the subtests or subsections of a language test, although this was conventionally associated with (construct) validity, not reliability (see Van der Walt & Steyn 2007: 196; Pot & Weideman 2015: 34).

measurement to political and ethical considerations and aspects such as transparency and accountability. However, as Van Dyk argues, the responsible design of assessments, and by implication also of the HL papers, demands more than awareness of the social consequences of testing; there is a need to be able to justify the foundational aspects born out in the validity of the construct and the reliability of the measurement. Evidence of both a theoretical and empirical nature is still necessary for the quality of assessment to be demonstrated convincingly.

In the instance of the present study, new sets of examination papers are designed each year and no use is made of psychometric measurement techniques. Although there have been calls in the reports made to Umalusi for item banking of refined examination items, this has not materialised yet. This may be a serious deficiency in the South African system, since standardised testing that includes the calculation of reliability coefficients and/or the use of data-banked items with high productivity is commonly used elsewhere to ensure comparable levels of difficulty across different examination papers when measuring the achievement of school leavers (see Zucker 2003 for a fuller exposition). It is surprising, therefore, that though the means and technical resources to do so are readily available, there has not been any movement in this direction.

The consequences of unreliable measurement are obvious. Even if the same test or examination paper were given to a group of candidates, without acceptable reliability estimates there could be little chance of ensuring equivalence or eventual fairness of assessment. However, in view of the enormous size of the Grade 12 HL examination population and the vast numbers of markers needed[16], it is debatable whether Umalusi would consider employing statistical means to correlate the marking of examination scripts. It is further unlikely that individual examination scripts would be marked by more than one person to improve the reliability of scores. Possible ways to enhance the reliability of the HL papers will be suggested in Chapter 6, which examines alternative formats to the current structure of the examination papers.

[16] In December 2014, around 7 million examination scripts in the respective school subjects were marked by a team of 41 564 markers at 117 marking centres located across the country (Department of Basic Education 2014). An average of approximately 472 348 learners write HL papers 1-3 each year (Du Plessis & Du Plessis 2015: 5), and this number rises every year; in 2015 it topped 500 000.

3.4.4 The notion of practicality

In as much as test designers seek to achieve high reliability and validity in their measuring instruments, they are constrained by practical limitations. Lado's (1961: 31) question of several decades earlier, "Does the test measure what we want it to test in a reasonable time considering the testing situation? If it does, the test is practical and economical", is still relevant. Bachman and Palmer (1996:9) agree that practicality is an important element that impacts on the usefulness of a test and include it in their list of six quality control variables.[17] Green (2014: 58) sees practicality as a "necessary condition for all assessment" and emphasises its importance by considering it to be the first of four qualities of "effective assessment systems":

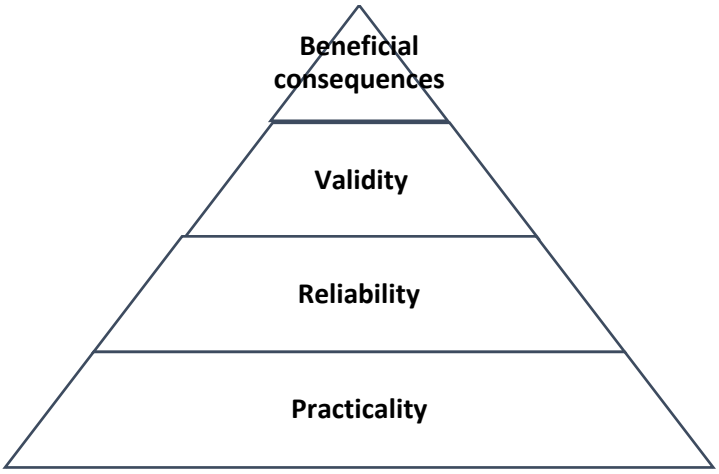


Figure 3.3: Four qualities of useful assessments (Green 2014: 58)

By placing practicality at the base of the cone, Green is suggesting that this should be the point of departure when designing a test; after all, there are definite limits as to what can be tested in a timed examination and how this can be executed in a feasible manner. Nonetheless, it would not be advisable to proceed with test design purely on the basis of practicalities. Weir (2005: 49) admonishes that the "convenience of the method should not be allowed to subvert the measurement of the construct". It would thus be irresponsible to select a convenient method of testing if the latter interfered with the generation of validity evidence.

[17] The other five variables are reliability, construct validity, authenticity, interactiveness (all discussed earlier in this chapter) and impact (i.e. social consequences of a test).

By placing beneficial consequences at the apex of his diagram of important test qualities, Green (2014: 58) wishes to illustrate that the "ultimate objective of any assessment system" is to benefit test users and test takers. However, any positive outcome of testing should be considered to be a desirable test effect. It may be argued that a test is more likely to have beneficial consequences if its design incorporates principles such as validity, reliability and practicality. One important consideration is whether the element of practicality will lead to positive washback after the testing in terms of improved teaching practice and skills development, an aspect referred to as consequential or systemic validity (Frederiksen & Collins 1989: 27).

It is clear that practicality has more than one dimension and that it refers to the relationship between the available and required resources necessary for the design, implementation and use of the test. In the framework of Weideman (2009a), it features under the regulative side of language testing and straddles social and economic considerations.

Table 3.4: Constitutive and regulative moments in applied linguistic designs (Weideman 2017: 224)

Applied linguistic Design	Aspect/function/ dimension/mode of experience	Kind of function	Retrocipatory/anticipatory moment
Is founded upon	Numerical	Constitutive	Systematicity
	Spatial		Limits, range
	Kinematic		Internal consistency (technical reliability)
	Physical		Internal effect/force (validity)
	Biotic		Differentiation
	Sensitive		Intuitive appeal (face validity)
	Analytical	Founding	Design rationale
Is qualified by	Technical	Qualifying/leading function (of the design)	
Is disclosed by	Lingual	Regulative	Articulation of design in a blueprint/curriculum/plan
	Social		Implementation/ administration
	Economic		Technical utility, frugality
	Aesthetic		Harmonisation of conflicts, resolving misalignment
	Juridical		Transparency, defensibility, fairness, legitimacy
	Ethical		Accountability, care, service
	Faith		Reputability and trust

In the above paradigm, the administration of the test technically links the assessment instrument to a social context, which may preclude the assessment of certain abilities. For example, owing to logistical constraints, oral ability forms part of school-based continuous assessment that is carried out by teachers in the classroom, unlike the assessment of other abilities in HL papers 1-3, which is carried out at examination centres and scored anonymously by teams of appointed examiners. In addition to social aspects, the technical function of a test is also analogically connected to the economic dimension in terms of utility or frugality, which pertains to resource and financial considerations, but also includes Bachman and Palmer's (1996) notion of "usefulness" referred to above. As part of the validation process, certain trade-offs have to be made between reliability and utility. For example, while longer tests tend to produce a higher reliability coefficient (Geldenhuys 2007), this may prove to be too demanding on the available resources, or

the time needed to administer them. Green (2014) relates this to efficiency in language assessment, and the trade-off between costs and benefits:

> ... the extent to which the commitment of resources to a system of assessment is justified by the benefits it brings. As in other areas of expenditure, assessment choices involve a complex balancing of benefits and costs. (Green 2014: 60)

If we consider the lack of credibility in the results of the exit-level examination mentioned in Chapter 1, it is questionable whether the current systems of assessment in Grade 12 should be allowed to continue. Can a case be made "that the benefits of maintaining them outweigh the costs" (Green 2014: 60)? Should a means not be devised to reduce the marking-intensive nature of the HL papers, for example, and increase the reliability of the results at the same time? Such a modification would in fact be highly useful, both in terms of frugality in respect of resources employed, and as regards gaining more reliable results that can be more usefully interpreted, and acquired with greater efficiency.

3.5 How the essential test qualities work together as a systemic framework

The increasingly sophisticated empirical indicators available in language assessment today make it more professional and increasingly specialised, necessitating the use of a comprehensive theoretical framework for test validation purposes. Irrespective of the different formulations of validity and reliability alluded to earlier in this chapter, the confidence that may be placed in any language test or examination is considered to be directly proportional to the multiple sets of evidence of ability collected in the process to support the validity of the evaluation instrument (Davies *et al.* 1999: 220, Van der Walt 2012: 145). By drawing on the theoretical and empirical indicators that feature in the approaches to validate a language test proposed by Bachman and Palmer (1996), Weir (2005), Weideman (2007; 2009a; 2017) and Green (2014) – discussed in detail in the preceding section – a comprehensive theoretical framework can be produced that reflects the main concerns of those accountable for the responsible design and administration of the HL examination papers.

Table 3.5: A framework for the responsible and principled assessment of HL ability based on the framework of Weideman (2017) and incorporating ideas from Bachman and Palmer (1996), Weir (2005) and Green (2014)

		Assessment timeframe			
		A priori evidence	**Per administratio evidence**	**A posteriori evidence**	
Theoretical justification for usefulness of HL examination papers and generalisability of results	**Constitutive concepts (foundational validity)**	**Analytical** Construct validity Articulated construct Design rationale		**Kinematic** Reliability Consistency of measurement	
		Biotic Content and context validity Authenticity and interactiveness Differentiation		**Numerical** Criterion-related validity Systematicity	
		Spatial Scoring validity Limits, range of measurement		**Physical** Internal effect	
		Sensitive Face validity/appeal			
	Regulative ideas (disclosed technical validity)	**Lingual mode** Design articulation Interpretability and meaningfulness of results	**Social mode** Conduciveness of organisational and institutional context for testing purposes	**Juridical mode** Use and impact of test scores Fairness Transparency	**Consequential validity**
		Aesthetic mode Alignment/harmonisation of instruction and assessment		**Ethical mode** Systemic validity/washback effect Accountability towards test users Care/beneficence	
		Economic mode Usefulness in terms of resources		**Faith mode** Credibility Trustworthiness	

It is clear from the framework that there are both objective and subjective variables. Weideman's notion of objective validity referred to earlier is to be discerned in the constitutive concepts pertaining to the validity of the test blueprint generated as a measurement artefact (i.e. its objective validity), while the ability of the instrument to generate reliable (objective) scores on the basis of which subjective inferences of ability can be made and generalised to non-testing domains shows how objective measurements (scores) are related to subjective technical interpretation in the process of validation. The notions of validity and reliability that form an integral part of the constitutive dimension of the framework (foundational validity) are underpinned by theories related to language learning and language testing. The respective modes that are included amongst the regulative aspects pertain more to issues of a disclosed technical validity and the socially contextualised employment of tests.

The framework furthermore illustrates how the normative moments "condition the design of an applied linguistic artefact" (Myburgh 2015) such as the HL examination papers to ensure that validity evidence is obtained systematically in multiple ways. An aspect not discussed in the preceding sections on validity is that of face validity:

> The degree to which a test appears to measure the knowledge or abilities it claims to measure, as judged by an untrained observer (such as the candidate taking the test or the institution which plans to administer it). (Davies *et al.* 1999: 59)

An alternative explanation of face validity is the degree to which a test meets the expectations of its users and the test takers and how acceptable the test is deemed to be by its stakeholders (McNamara 2000: 133). In this regard, apart from being considered part of the sensitive dimension of experience, face validity could also be construed as a technically disclosed faith dimension that forms part of consequential validity. The HL papers may thus have initial face validity amongst those who design them and oversee the HL examination, but this view will not necessarily be shared by other stakeholders after the administration of the examination. For the examination to become credible and trustworthy it will have to gain more than face validity.

Table 3.5 shows the analogical relation between the technical and the other modes of experience and how these relate to language testing principles such as validity and reliability. The articulation of a test blueprint is indicative, for example, of the analogical

relation between the technical and the lingual mode of experience. The interpretability and meaningfulness of test results depends on the alignment of the dimensions. Weideman (2007: 601) points out that in attempting to align all of the mentioned factors, the test designer "brings them into harmony within the design", which in turn refers to the aesthetic dimension.

The social context within which the testing takes place has hitherto not received much attention. Apart from the inauthenticity of assessing language use in an artificial test setting, there are also likely to be major contrasts in the institutional and organisational environments in which tests or examinations are written. Less well-resourced schools may not have adequate lighting and facilities, for example, which can introduce a measure of inequality into the assessment process. Similarly, students in poor areas may not have the same benefits of good nutrition to enable them to perform optimally mentally and physically, as would be the case with students residing in more affluent areas.

The juridical dimension has a bearing on the public defensibility and fairness of a test, or what is usually referred to as accountability. The way that a test or examination can be used fairly as a gateway to higher education or the workplace serves as an example hereof. The ethical domain in turn elicits considerations of beneficence, such as how administering the HL examination can contribute towards better teaching practice and the care or development of examinees. Faith as a regulative dimension reflects belief in a system on the basis of its reputed value, credibility and trustworthiness. These three regulative conditions contribute towards the notion of consequential validity, which refers to both the intended and the unintended and potentially harmful effects of tests (Messick 1980, 1981; Shohamy 2001a, 2001b, 2006; Fulcher & Davidson 2007; Xi 2010 and Young 2012). As Fulcher 2010: 17 points out, "tests have been used for as long as we can tell as tools to control teachers and educational systems to deliver the kind of society and economy envisioned by the powerful." Although tests clearly can have negative consequences (especially where they result in narrowing of the curriculum), primarily they should be seen in a positive light as a means of ensuring that language learning has taken place and that learners are being equipped to participate fully in society as articulate citizens. The beneficent part of testing should thus be emphasised and planned for. By acceding to the foundational constitutive conditions of language testing and by incorporating different forms of assessment into the school language programme,

negative washback can potentially be averted and learners can be assisted to develop their language knowledge and proficiency as broadly as possible. The issue is: the credibility and trustworthiness of the current high-stakes examination for HLs must be restored – the analogical link between the technical qualification of their design and the dimension of faith yields a regulative condition that must be fulfilled. Moreover, without fairness, another regulative condition, their credibility will continue to be undermined.

Regulative principles are just as important as constitutive conditions, but are more likely to have positive consequences if the foundational conditions are met. As pointed out, the reciprocal relationship between the foundational (constitutive) and technical (regulative) aspects provides the theoretical justification for the examination (Weideman 2009a: 246). However, notwithstanding the contribution of all the mentioned modes towards the responsible design of the HL papers, the emphasis of the current study will fall on the generation of *a priori* and *a posteriori* evidence of a constitutive nature, since this is considered essential for the purposes of analysing the content of a selection of HL papers and for the development of a theoretical rationale for greater equivalence of standard.

3.6 Conclusion

This chapter started by placing the current study within the theoretical framework of applied linguistics as a discipline of design. Alternative paradigms of thinking were discussed to show how views of language have changed over the last approximately six decades and how these have been reflected in language teaching and testing. Of particular significance is the fact that no scientific – in the sense of authoritative or conclusive – method exists to guarantee solutions to language problems. Attempts to implement purportedly scientific methods to teaching languages have proven to be inadequate and future views in the field of language teaching are likely to reflect combinations of earlier ideas owing to the continuity that characterises the design of applied linguistic interventions. Furthermore, it was pointed out that language assessment has both an objective and subjective component and that the dynamic and complex nature of language precludes any singular best method of testing. Nevertheless, designers of language tests carry a responsibility to ensure that the artefacts they develop comply with certain principles that are accepted internationally by the language testing fraternity in the interests of preventing harmful and unintended consequences of testing.

Three essential qualities for the responsible design of language tests that could potentially contribute evidence for the construction of what is conventionally known as a validity argument have also been discussed above. The objective here was to generate a framework that could assist the HL examiners to design papers that reflect current theories and applications of language and that facilitate more equitable and socially responsible forms of language assessment. A multi-faceted approach was adopted rather than a conflated unitary view of construct validity as the superordinate form of all other validities.

From the discussion in this chapter of language testing principles, it is apparent that the process to validate a language test or HL examination is both systematic and systemic. All constitutive and regulative conditions for responsible test design need to be complied with in a coordinated way. A shortcoming in any one of the applied linguistic design tenets can undermine the usefulness of the measurement system as a whole. Through paying attention to the way that a number of assessment concepts cohere, we can make great strides towards responsible language testing.

This chapter has covered principles applicable to the design of the HL papers. However, it should be noted that language teaching and testing are two sides of the same coin. The curriculum on which the language papers are premised will thus be discussed in the chapter to follow as a basis for achieving conceptual clarity on the superordinate construct of the Grade 12 HL examination.

4 Conceptual clarity on the underlying construct of the HL papers

4.1 Introduction

In view of the fact that the Grade 12 HL examination is not only an assessment of language proficiency, but also the measurement of mastery of the content of the curriculum that defines the high-level proficiency required by the syllabus, conceptual clarity should in the first instance be gained on the underlying construct of the examination papers as encapsulated in the aims and principles of the curriculum and learning programme. Subsequent to this, the construct can be articulated in a number of specifications of ability, also deriving from the curriculum, for the purpose of developing a system of categorisation and a framework for the analysis of the examination papers. At this point it should be mentioned that the examination papers scrutinised in the current study were based on the previous curriculum, the National Curriculum Statements (NCS). A cursory comparison of the new and old curricula, however, reveals that the new curriculum is essentially an abridged and more user-friendly version of its predecessor, and that the core objectives remain the same. A decision has thus been taken to examine the extent to which the HL papers are aligned with the new CAPS document, and whether these should be retained in their current format once the new curriculum has been rolled out in full.[18]

Attempting to problematise a unitary construct for measuring the language ability of Grade 12 learners is somewhat ambitious in light of the multi-faceted nature of language and the reality that more than one ability is likely to be assessed simultaneously. As a first step to address the construct issue, this chapter examines the new Curriculum and Assessment Policy Statement (CAPS) from the perspective of the aims of the new curriculum, theories of language learning that inform assessment practices and those language abilities reflected in the policy document. Hereby conceptual clarity can be achieved on the underlying construct before proceeding to apply the theoretical framework aimed at ensuring greater equivalence of standard and construct across the respective HLs.

[18] This process was scheduled to be completed in 2014.

4.2 General aims of the South African Curriculum

In terms of the provisions of section 61 of the South African Schools Act (Act No. 84 of 1996 as amended, RSA 1996b), the Minister of Basic Education has the right to decide on the minimum outcomes and standards of the language curriculum, as well as determine the processes and procedures for the assessment of learning in all South African schools. Any conceptual framework for the assessment of the HLs would thus need to be developed within the parameters of the norms and standards of the National Curriculum Statement Grades R-12, to be implemented in full by 2014. For the purposes of clarity and unnecessary avoidance of repetition of detail, all references in this thesis pertaining to the National Curriculum Statement Grades R-12 shall be understood to be referring to the Curriculum and Assessment Policy Statement (CAPS) for the Further Education and Training Phase (Grades 10-12) and HLs in particular.

A distinction is made in CAPS between different levels of teaching that apply to each of the eleven official languages and other non-official languages on offer. These levels are referred to as Home Language (HL) level and First Additional Language (FAL) level. Technically speaking, the notion of HL would refer to that language used by a learner within the context of the home and concurrently the first language acquired. In reality, however, a learner may have been exposed to more than one language in the home from infancy and the language spoken in the home context may not be offered as a subject in all schools, thus necessitating the learning of an Additional Language at HL level. In order to resolve this dilemma, the CAPS document makes it clear that the distinction of HL applies to the *level* at which the language is offered rather than the language itself. The standard thus set for HL level is higher than that set for FAL level, although in pragmatic terms the competency level of a learner may be the same for both levels. In light then of the level of difficulty that is to distinguish HL from FAL and on the basis of the information contained in Section 2.1 of CAPS, two levels of proficiency can be identified which are applicable to the assessment of language at HL level:

1. Social level:
   "the mastery of basic interpersonal communication skills required in social situations";
2. Academic/Educational level:
   "cognitive academic skills essential for learning across the curriculum", including "literary, aesthetic and imaginative ability". (Department of Basic Education 2011a: 8)

The two levels of mastery referred to above derive from the work of Cummins (Cummins & Davison 2007: 353) who noted distinctions in the proficiency of first language and second language speakers in bilingual educational settings, and the differences in ability required for "*basic interpersonal communicative skills (BICS)* or conversational language and *cognitive academic language proficiency (CALP)*, academic language". A much higher order of thinking and level of language ability are needed for academic purposes than social settings. It is this more advanced ability that is to be reflected more prominently than basic conversational ability in the HL component of the school language programme, if learners are to participate in "society as citizens of a free country", have "access to higher education", and be able to make the transition from "education institutions to the workplace" (Department of Basic Education 2011a: 4).

Reference is also made at the end of Section 2.1 to the necessity of being able to use an additional language at a sufficiently high standard in order to be able to gain access to "further or higher education or the world of work" (Department of Basic Education 2011a: 9). In view of the latter, a third level of proficiency can then be added to the two mentioned above, that of the economic/occupational level which would require the mastery of language skills needed for professional and employment purposes and access to trade and industry. The three levels identified above – social, academic/educational and economic/occupational – are operationalised in CAPS in a number of different fields of discourse, which will be dealt with later.

Apart from the general aims of the curriculum, CAPS identifies the following specific aims for the learning of languages:

> Learning a language should enable learners to:
>
> - acquire the language skills required for academic learning across the curriculum;
> - listen, speak, read/view and write/present the language with confidence and enjoyment. These skills and attitudes form the basis for life-long learning;
> - use language appropriately, taking into account audience, purpose and context;
> - express and justify, orally and in writing, their own ideas, views and emotions confidently in order to become independent and analytical thinkers;
> - use language and their imagination to find out more about themselves and the world around them. This will enable them to express their experiences and findings about the world orally and in writing.
> - use language to access and manage information for learning across the curriculum and in a wide range of other contexts. Information literacy is a vital skill in the 'information age' and forms the basis for life-long learning; and

- use language as a means for critical and creative thinking; for expressing their opinions on ethical issues and values; for interacting critically with a wide range of texts; for challenging the perspectives, values and power relations embedded in texts; and for reading texts for various purposes, such as enjoyment, research, and critique. (Department of Basic Education 2011a: 9)

Together these objectives provide the background against which language learning and assessment should take place and are of particular relevance when deciding on appropriate language-related tasks that learners should be able to execute, if they are to be able to operate as *highly* literate citizens. CAPS underwrites the principle of "*high* knowledge and *high* skills" and the minimum standards to be attained are to be "*high*, achievable standards in all subjects" (Department of Basic Education 2011a: 4). Any undermining of this level of ability in the current HL papers will negatively impact on the validity of the exit-level examination.

A detailed exposition of the content that is to be covered in the language classrooms so as to meet these specific aims and how the needed skills are to be developed is provided in more detail in Section 3 of CAPS. Together the respective sections provide a full articulation of abilities to be mastered in the Further Education and Training (FET) Phase and constitute the basis for formulating a general construct for the assessment of HLs. It is immediately clear, however, that the conceptual terminology adopted here articulates the intention to develop in learners a differentiated language ability so that by the end of their school careers they have mastery of language(s) in a wide range of different (educational and academic; aesthetic; political; economic; social and informational; ethical) contexts and situations. It is equally important, moreover, to acknowledge that the same starting points and assumptions that constitute the basis for the teaching of the language curriculum need to inform the assessment side of language learning too.

The policy statement therefore clearly emphasises the provision of language teaching that is relevant to different realms and applicable at different levels. Both functional levels and a differentiated variety of language can be identified, not only in the instance of HL teaching, but also in terms of FAL teaching. No doubt the rationale for this is to be found in the multicultural context of the South African classroom and multiple identities of the learners themselves, realities that complicate the identification of what is to be understood as the HL or FAL of a learner. In summary, from the general aims and principles of the policy document, different lingual realities or spheres can be identified which form part

of a contextual and theoretical framework within which language teaching and assessment should take place.

4.3 Conceptual distinctions that inform CAPS and may serve as a basis for the formulation of an underlying construct for language assessment

There is no doubt that the conceptual framework that underlies CAPS goes back to linguistic ideas originating in the early 1970s on a differentiated communicative competence (Habermas 1970; Hymes 1972; Halliday 1978) that supports actual language use by varied repertoires of functionally defined language acts (Searle 1969; Wilkins 1976). In their subsequent development, these constitute socially informed ideas about language that have not only disclosed and broadened our perspective of what constitute language ability and language use – that mastery of language, for example, is much more than having a grammatical command of it – but have also weathered well. They have informed, for example, cutting-edge language teaching in Australia, and have provided the theoretical rationale for a whole spectrum of genre-based approaches to language teaching (Wyatt-Smith 1997; Hyland 2003; Carstens 2009). It should therefore not be surprising that CAPS refers to the teaching approaches underlying it as communicative and text-based (Department of Basic Education 2011a: 11). Further to the above, language structures are to be learned in an integrated way in the school classroom so that they "aid successful communication" and are "linked to the functional uses of language in different social settings" (*ibid.* 2011a: 10-11). This may support the design of integrated language examinations, rather than the current separation of abilities encountered in Papers 1 and 3.

At the basis of these disclosed and enriched sociolinguistic ideas about language is the notion that it operates in particular contexts and lingual spheres, relating to what Halliday refers to as fields of discourse (Halliday 1978: 221). Weideman (2009b: 40) explains that these spheres may be considered material since they are governed by "*typical* norms and principles that give a different content to the factual language used" within a situation. Consequently, distinct lexical and syntactic differences can be discerned in the language used in diverse contexts. The norms that regulate the lingual spheres are typical because they apply to social forms or relationships that require a typical type of language usage

within a particular temporal and structural context. Human beings as the users of language fulfil a subjective lingual role in which they then select from a repertoire of already developed registers in order to produce a lingual object deemed appropriate to a given situation. Most of the linguistic choices made are unconscious ones (Bloor & Bloor 2013: 4), but this does not guarantee their suitability.

Learning appropriate lingual responses forms part of what is referred to as "language socialisation", a term coined by linguistic anthropologists Elinor Ochs and Bambi Schieffelin in the 1980s to foreground the intricate relationship between language learning and the process of socialisation (Riley 2010). According to this notion, the two processes of socialisation and the development of communicative competence are mutually dependent. Socialisation is viewed as a "dynamic and language-rich process, and acquisition of communicative competence as a culturally coded experience" (*ibid.*: 399). In as much as socialisation is made possible through language, language use and norms for that use are acquired through socialisation. This accentuates the importance of exposing learners to culture-specific contexts in which socialisation and knowledge construction take place, both inside and outside the classroom, if they are to learn to make the appropriate choices that signify linguistic and communicative competence.

Weideman elucidates the above notion further by stating that materially distinct lingual spheres are indicative of a "differentiated classification of language types that is inextricably bound up with the *subjective* human lingual capacity for producing *objective*, factual language in various social spheres" (Weideman 2009b: 41). Hence the distinction needs to be made between lingual fact and lingual norm. The factual context alone is inadequate when determining what type of language should be used, because of the "normative principles of a logical, aesthetic, legal, technico-formative, economic, social, ethical or confessional nature" (*ibid*. 2009b: 41) that apply in different types of discourse: academic, legal, social, political, and so on. One thus needs to distinguish between the language situation itself, and the conditions for using language in that situation. Such a distinction is extremely important when articulating any construct or ability to be assessed within the context of a language examination and obviously also when designing assessment rubrics.

Weideman (2009b: 48) points out that "material or typical differences are discernible too on almost every level of language: phonological, lexical, syntactic, semantic, and so forth". The use of different dialects within communities and varying tones of voice to convey meaning further illustrate this. Language derives its meaning from more than a code or set of symbols. When using the term "typical" to demote lingual spheres, this should be understood as referring to that which is "lingually typical" (*ibid.* 2009b: 49).

Humans seem to have an inherent ability (a communicative competence) to recognise different varieties of language. Vocabulary plays a role in distinguishing between material lingual spheres, but is insufficient on its own because language is qualified by different experiential modalities. As such it would seem that each sphere has a typical language of its own. Moreover, the social structure in which the language is employed could be responsible for further distinctions.

The advantage of employing a systematic and theoretical framework such as the above one pertaining to material lingual spheres when teaching and assessing language is that provision can be made in the school examination system for both a differentiated language ability within different material lingual spheres, as well as for a generic language ability incorporating functional and formal aspects of language. This would also allow for the accommodation of attributes peculiar to a particular language group and possible differences in the maturity and status of some of the indigenous languages which may not be equally well represented or resourced in respect of all of the material lingual spheres.

4.3.1 Fields of discourse identified in CAPS

In terms of the content of the language curriculum outlined in CAPS and discussed above, the dominant material lingual spheres of relevance to the teaching and assessment of the respective languages would seem to be the following:

- social (including inter-personal communication and the handling of information)
- academic/educational (including academic and scientific language and advanced language ability)
- aesthetic (including literature and art)
- economic/occupational (including the world of work and commerce)
- ethical (including an appreciation of the values embedded in language use)
- political (including the critical discernment of power relations in discourse)

The only material lingual sphere not as prominently reflected in CAPS as some of the others, and thus perhaps not as pertinent to learners at school level, seems to be the legal or juridical sphere, which may be considered to be of too specialised or specific a nature. Incidentally the first four dominant material lingual spheres identified in CAPS correspond closely with the list of approved teaching subjects stipulated in Annexure B of the policy document, *National policy pertaining to the programme and promotion requirements of the National Curriculum Statement Grades R-12* (Department of Basic Education 2015b), and listed in Table 4.1.

Table 4.1: Lingual spheres and approved school subjects that comply with the programme requirements of Grades 10-12

Type of sphere	Subject
Social	Human and social studies, languages
Aesthetic	Culture and arts, languages
Academic/educational	Mathematical, physical, computer and life sciences, agriculture, engineering and technology, languages
Economic/occupational	Business, commerce and management studies, consumer, hospitality and tourism services

The sociolinguistic ideas referred to in this chapter generally make a distinction possible between the norms for language that are provided by and in such lingual spheres or discourse types, and the factual language usage ("texts") that occur in the various spheres of discourse.

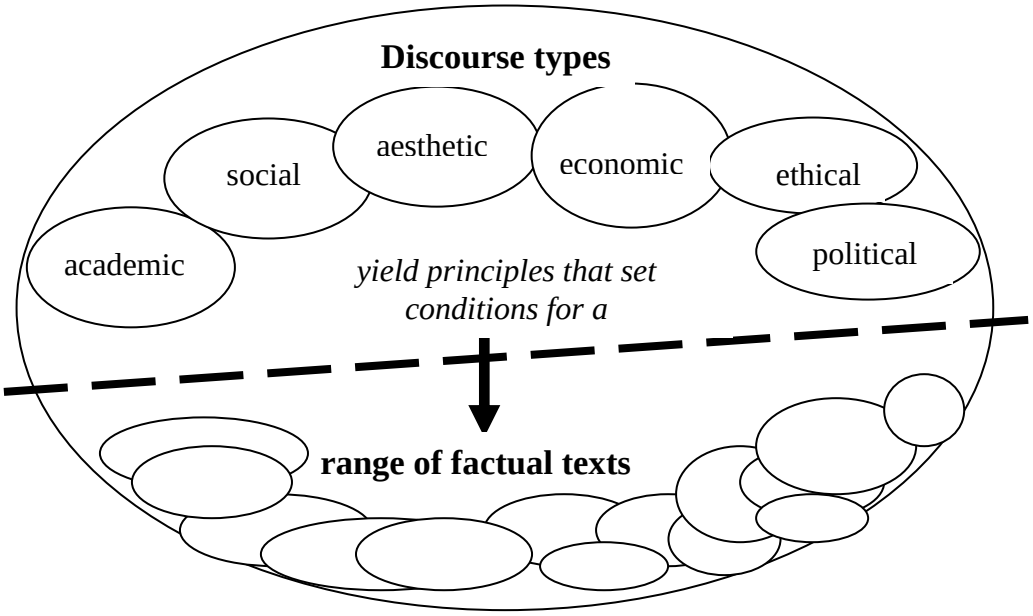


Figure 4.1: Conditions for language

Discourse types therefore provide the conditions or requirements for a wide range of factual texts. This differentiated variety of discourse types therefore supports the notion not only of a differential ability of language use (a differentiated communicative competence), but also guarantees the differences in different kinds of text. CAPS makes provision for this in enumerating a wide variety of text types to which learners should be exposed, but does not seem to take into account to what extent each of the indigenous languages has developed historically and that not all discourse types may be applicable to all languages at this stage. The combinations of texts relating to the teaching of reading and writing and relevant to the HL examination papers[19] are summarised in Tables 4.2 and 4.3 below according to the dominant spheres. Ethical and political language may occur in any of these.

Table 4.2: Fields of discourse illustrating factual reading texts in CAPS

Type of discourse	Type of factual reading text
Social	Letters, diaries, invitations, emails, sms's, twitter microblogs, notes, reports, telephone directories, television guides, dialogues, blogs, Facebook and social networks entries
Aesthetic	Novels, dramas, short stories, poetry, films, radio and television, series/documentaries, radio dramas, essays, biographies, autobiographies, folk tales, myths and legends, songs, jokes, photographs, illustrations, music videos, cartoons, comic strips
Educational/Academic	Dictionaries, encyclopaedias, schedules, textbooks, thesauruses, timetables, magazine articles, newspaper articles, editorials, notices, obituaries, reviews, brochures, speeches, charts, maps, graphs, tables, pie charts, mind-maps, diagrams, posters, flyers, pamphlets, signs and symbols, television documentaries, internet sites, data projection, transparencies, caricatures, graffiti
Economic/Occupational	Formal letters, minutes and agendas, advertisements, web pages

How all the HLs cope with the requirement to secure mastery in all these different kinds of discourse texts by including them in the language instruction of the classroom, and specifically in the instructional material (textbooks) provided, is beyond the scope of this study, but would certainly answer questions about which kinds of discourse are available in these different languages.

[19] The assessment of listening and speaking proficiency forms part of school-based assessment and as such falls beyond the scope of the present study.

Table 4.3: Fields of discourse illustrating differentiated writing abilities in CAPS

Type of discourse	Type of factual text to be written
Social	Formal and informal letters, dialogues, speeches, interviews, obituaries,
Aesthetic	Narrative and descriptive essays, reviews of art, films or books
Educational/Academic	Literary essays, argumentative, discursive and reflective essays, reports, newspaper articles, magazine articles
Economic/Occupational	Transactional texts, formal letters, minutes, memoranda and agendas, interviews, curriculum vitae

Some overlapping of fields or spheres of discourse is naturally possible. For example, a magazine article may be both aesthetic and educational, while a formal letter could apply both in the economic and social realms. The ability of the Grade 12 learners to operate at different levels of proficiency in these diverse spheres and to display versatility in terms of register and style is what needs to be assessed summatively in the final exit-level examinations. The only way to do this is through examination specifications in the form of language-related tasks selected on a systematic basis so as to be able to provide sufficient evidence of ability. Whether the current structure of the examination papers and selection of tasks may be considered to provide an adequate basis for evaluating language ability remains to be seen; this is a question that will be addressed in the detailed analysis of the language papers. An underrepresentation of essential abilities or the inclusion of unessential or irrelevant tasks would undermine validity (see Weir 2005: 18). Some tasks may only be applicable to a small category of individuals in society, for example creative writing ability commensurate with that of a novelist or poet (i.e. artistic ability). A case in question would be the kinds of tasks included in Section A of Paper 3 (Writing). A further concern would be to what extent all of the text types would apply to or be available in those languages that have not been developed to the level of language of instruction at an institution of higher education.

4.3.2 Generic and differentiated language abilities

Sections 3.2 and 3.3 of CAPS cover what learners should be able to do in terms of the reading and viewing process, as well as when writing and presenting (Department of Basic Education 2011a: 22-40). These detailed abilities should be considered the full set

of test specifications from which a selection has to be made when articulating the blueprint for the examination of high language ability.

Generic language abilities refer to the kinds of abilities needed to access and comprehend written or printed materials to the extent that questions and tasks based on the given texts can be answered without the requirement of displaying differential ability. In other words, marks are not allocated for the style or construction of the written response, or the writing prowess displayed. No adjustment of register or tone, for example, is needed. Generic abilities specified in CAPS include reading and interpreting texts, knowledge of vocabulary, language structures and conventions, and text organisation.

4.3.2.1 Reading and viewing for comprehending and appreciating texts

Four broad categories are specified: i) the reading process; ii) interpretation of visual texts; (iii) vocabulary development and language use; (iv) sentence structures and the organisation of texts. The detailed lists of tasks to be performed are provided below in the order that they occur in CAPS.

(i) The reading process (applicable to all text types)

A three-phase process is followed, which includes pre-reading, reading and post-reading strategies requiring the listed abilities (here provided in the order that CAPS gives):

Skim and scan text features and book parts
Make predictions
Work out the meaning of unfamiliar words and images
Make sense of the text
Make connections
Monitor comprehension
Ask and answer questions
Visualise
Infer
Read for main ideas
Attend to word choice and language structures
Use structure and language features to recognise text type
Make notes
Summarise main and supporting ideas[20]
Compare and contrast
Synthesise
Evaluate
Draw conclusions

[20] These and some of the following abilities in the list are actually writing tasks that require comprehension of a text passage.

Express own opinion
Reproduce the genre in own writing[21]
Distinguish between fact and opinion (critical language awareness)

(ii) Interpretation of graphic and visual texts

In addition to the generic abilities above, the following additional tasks are applicable to graphic and visual texts:

Examine how layout is a key aspect of popular websites
Examine how advertisers get attention
Examine how movement and colour play key roles in persuading the reader to move to other sites
Understanding how language and images reflect and shape values and attitudes
Identifying images that are sexist, racist, ageist or stereotyped
Investigate the impact of use of font types and sizes, headings and captions
Analyse and respond to cartoons/comic strips

Many of the tasks in Section (ii) are not language-related *per se* and may be considered to fall within the discipline of Communication Science or Visual Semiotics. The desirability of including such task specifications in language-specific examinations is debatable and will be deliberated in the ensuing chapter as part of the analysis and evaluation of abilities assessed.

(iii) Vocabulary development and language use

In line with communicative language teaching all vocabulary items and aspects of language use are to be embedded in authentic texts and the emphasis should be on the reasons underlying the choice of words and figures of speech rather than merely identifying these. CAPS lists the following:

Identify and explain the use of figurative language and rhetorical devices
Distinguish between denotation and connotation
Determine the meaning, spelling, pronunciation, syllabication and part of speech of unfamiliar words using reference books
Identify the meaning of common prefixes and suffixes
Use knowledge of prefixes, suffixes and common roots to determine the meaning of words and their connections to word families
Use textual context and cues to determine the meaning of unfamiliar words
Distinguish between multiple-meaning words in uncomplicated texts about concrete topics[22]
Recognise common allusions[23]
Demonstrate an understanding of common phrases, proverbs and idiomatic language

[21] This is a writing task that requires more than generic ability to read and understand a text and should, if there is a strict skills-based distinction to be made, not be listed here.
[22] Exactly what is meant by "uncomplicated" and "concrete" is not clear, nor the reason for the specification of texts. Surely polysemes occur in a variety of texts on all kinds of topics of varying degrees of difficulty.
[23] This may require additional knowledge and could be biased towards some learners.

Evaluate how words from various cultural origins have an impact on text[24]
Distinguish between commonly confused words: homophones, homonyms, homographs, synonyms
Retell a story or sentence using different words (synonyms and antonyms)[25]
Use one word for a phrase
Use collocations

(iv) Sentence structures and the organisation of texts

Although these are listed under Reading and Viewing, CAPS advises incorporating such tasks in the "writing lessons" (p. 24), i.e. ensuring that learners can employ the listed language structures and conventions in writing. Learners are required to identify the following and be able to explain their conventional use:

Transition words/conjunctions
Abbreviations and acronyms
Verb forms and auxiliaries expressing tense and mood
Simple, compound, complex, and compound complex sentences using clauses, phrases and conjunctions
Active and passive voice
Direct and indirect speech
Correct word order
Concord, articles, infinitives, copulatives, prepositions
Punctuation

In addition to the above, learners are expected to analyse text organisation and the role of transitional/signal words in texts that include the following aspects:

Chronological/sequential order
Explanation
Cause and effect
Procedure
Comparison/contrast
Order of importance
Spatial order
Choice paragraph
Classification paragraph
Description paragraph
Evaluation paragraph
Definition paragraph
Expositions
Reports
Concluding paragraph

The study of literature is included in CAPS as part of Section 3.2 on Reading and Viewing. Considering that this is a different subject field with a very specific vocabulary and content, and that there is a separate examination paper for literary studies, it may be

[24] This is specialised sociolinguistic knowledge and would prejudice certain learners.
[25] This would involve a form of writing or speaking.

preferable to make it a separate school subject in CAPS as well. (This idea will be motivated further in subsequent chapters.) Three main genres of literary texts are listed in the curriculum for detailed study, namely novel, drama and poetry. All the other genres are considered to be for enrichment purposes only. The current research project will focus specifically on the examination of Language in Context and Writing (Papers 1 and 3 respectively), but will venture to make some recommendations on the structure and weighting of Paper 2.

4.3.2.2 Writing and presenting

A process approach is adopted in which learners are required to demonstrate their ability by applying their knowledge of the different structures and features of a variety of text types, knowledge of sentence and paragraph conventions, as well as ability to use punctuation in a number of writing tasks (CAPS, Department of Basic Education 2011a: 30). Both generic and differentiated abilities are involved.

(i) Generic writing abilities

Apart from the planning, pre-writing, drafting and revising stages of writing, learners should display the following general abilities, which should form part of the assessment rubrics used to evaluate writing ability:

Use main and supporting ideas effectively from the planning process
Take into account purpose, audience, topic and genre
Use appropriate words, phrases and expressions so that the writing is clear and vivid
Display an identifiable voice and style in keeping with the purpose of the writing.
Demonstrate own point of view supported by values, beliefs and experiences
Use information from other texts to substantiate arguments
Write in such a way that there is no ambiguity of meaning, redundancy or inappropriate language
Use punctuation, spelling and grammar correctly
Use appropriate register, style and voice
Construct a variety of sentences of different lengths and complexity using parts of speech appropriately
Show knowledge of cohesive ties
Use active and passive voice
Use direct and indirect speech
Use affirmatives and negatives
Display knowledge of verbs, tenses and moods
Use interrogatives
Write different parts of a paragraph, including introductory, supporting and concluding sentences
Write different kinds of paragraphs (sequential, cause and effect, procedural, comparisons/contrasts, introductory and concluding paragraphs)
Write texts that are coherent using conjunctions and transitional words and phrases

(ii) Differential abilities

Learners are required to be able to produce a range of text types requiring particular formats and features within specific fields of discourse. Two broad categories of writing are mentioned: essays and transactional texts.

Types of essays that learners should be capable of writing include the following:

Narrative
Descriptive
Argumentative
Discursive
Reflective
Literary

The detailed specifications for each of the above will not be discussed at this point. It should, however, be noted that the desirability of including all of the above in the examination papers is to be questioned, since some essay types require much more than language knowledge and are specialised fields of writing in which other constructs are involved. Another preliminary concern is that learners should be able to apply the same stages of the writing process set out in the curriculum within an examination context, if there is to be alignment between the curriculum and assessment. Most essays require research and are not produced in a vacuum. This aspect will be examined in more detail as part of the content analysis of the writing components of the HL papers in Chapter 5.

The following transactional types of texts are listed:

Official and formal letters
Friendly and informal letters
Texts related to meetings (agendas, minutes, memoranda)[26]
Speeches, dialogues and interviews[27]
Formal and informal reports[28]
Reviews (books, films, etc.)[29]
Newspaper and magazine articles[30]
Curriculum vitae
Obituaries

[26] These should be based on authentic meetings if they are to be of any real value.
[27] CAPS specifies that these three kinds of texts should not be done in isolation as writing exercises.
[28] Artificial reports are to be avoided.
[29] These would require much study and advance preparation and would not be suitable for inclusion in a pressurised examination context.
[30] Research would be required.

Not all of these would be suitable for inclusion in an examination paper, especially where research or additional knowledge of a topic is needed, and some may be produced via electronic media necessitating the use of different formats and features. Of the mentioned texts, mainly the first two types (formal and informal letters) seem to be suitable for inclusion in an examination paper. The remainder should form part of school-based assessment where the different stages of process writing can be applied.

In summary, if one considers CAPS in its entirety, the assessment standards appear to be in order for HL level and the curriculum seems to be comprehensive. However, in as much as the curriculum and assessment standards may help to organise what should happen in the classroom, they provide no guarantee of contributing towards the quality of education or assessment practices, or of ensuring equivalence across different language examinations. Without denigrating the importance of standards and curricula, Davies points out that the emphasis needs to be shifted from setting standards to ensuring accountability, which he defines simply as a "way of explaining that what has been done is appropriate and necessary" (Davies 2010: 484). This aspect of accountability in language teaching and assessment is directly related to defining the underlying construct and articulating it on the basis of defensible theories of language and communicative competence.

4.4 Conclusion

From the afore-mentioned discussion of differentiated and generic abilities that the HL curriculum foregrounds it is evident that language is a complex phenomenon incorporating many and heterogeneous types of language, rather than a singularly identifiable object. Any language-related act occurs within a unique context that has a direct bearing on what kind of language is used. We can see that language has many levels, is dynamic and constantly changing.[31] This view of language is adequately reflected in CAPS both in the acknowledgement and representation of materially different lingual spheres, as well as the generic abilities reflected in the kinds of tasks that learners are expected to execute in the sections on functional language usage and formal language

[31] The field of English language teaching is currently moving in a new direction with growing recognition of its role as an international language. Socioculturally sensitive pedagogies are being foregrounded along with a new appraisal of what constitutes the variety of standard English. This could influence facets of the English HL curriculum in the foreseeable future.

structures and conventions. Taking all of the above into consideration, the general underlying construct for the HL examination papers has been conceptualised as follows and communicated in a report for Umalusi; that report states that we should be aiming at:

> … the assessment of a differentiated language ability in a number of discourse types involving typically different texts, and a generic ability incorporating task-based functional and formal aspects of language. (Du Plessis, Steyn & Weideman 2013: 20)

The above construct can be elucidated further on the basis of the distinction made in CAPS between three levels of mastery, namely basic communicative ability required in various social situations, professional proficiency for the workplace, and advanced ability for cognitive processing in educational and other contexts. It is the more advanced levels of proficiency that are to be reflected in the differentiated and generic abilities examined in the HL papers. Determining the desired standard of the examination paper can be seen as an integral part of the articulation of its construct, and any undermining of what is supposed to reflect a high ability will weaken the validity argument. Advanced ability is implied inherently in the conceptualised construct of CAPS in the differing ability required of examinees to respond to a variety of discourse types that demand the mastery of distinct language features.

Now that a conceptual framework has been provided and a superordinate construct been established, attention can be devoted to the articulation of this construct in a selection of language papers by examining the task specifications reflected in the papers and marking memoranda, so as to be able to express an opinion on whether these are sufficiently representative of the curriculum and whether fundamental principles in language testing are being applied.

The intention of the ensuing chapter is to analyse in full the November English HL papers covering the period 2008-2012 on the basis of the conceptual distinctions made in CAPS and accepted language testing principles applicable to the validation of language examinations. In this respect the analysis of the English papers will take precedence over the remaining official languages for three reasons. Firstly, the researcher's interest and competence lies in the field of English language studies, and secondly, English carries international status, a fact which necessitates ensuring that its assessment at school level in South Africa is on a comparable footing with global standards. A third reason that may

be cited is the fact that English is the dominant language of instruction at institutions of higher learning in South Africa, highlighting the importance of developing English language ability to a sufficiently high level for post-school qualifications. There is thus an added reason to ensure that the inferences based on scores obtained in the English HL papers will assist with the identification of students who have reached a level of language proficiency adequate for supporting academic learning at an advanced level. A brief comparative study will nevertheless be made of a selection of Afrikaans, English and Sotho HL papers in Chapter 6 to point out similarities or differences relevant to the topic of the thesis.

5 Content analysis of the Grade 12 English HL examination papers (2008-2012)

5.1 Introduction

The main purpose of this chapter is to determine what language abilities have been the focus of the exit-level examination, and whether there has been alignment between the examination and the language curriculum. This is an important aspect of theory-based validity, which in turn forms part of construct validity. As discussed in Chapter 3, the underlying language and cognitive processing that takes place when performing curriculum-related language learning tasks needs to be replicated during the language assessment event.

A further core objective of analysing the content of a selection of examination papers would be to determine whether the task specifications and kinds of examination items support the construction of a validity argument. Evidence of what the examinee knows and can do needs to be inferred from quantified responses provided to the selection of tasks and items. The way items are presented and marks allocated contributes towards the reliability of measurement, an essential prerequisite for the validation of the examination papers. Each examination component in the selection of papers under review will therefore be evaluated on the basis of the *a priori* constitutive conditions that pertain to the validity of construct, content and scoring. Although any discussion of scoring validity should include reliability coefficients, owing to the unavailability of statistical data it will only be possible to estimate the reliability of scoring, as pointed out earlier. Consideration will also be given in the ensuing discussion to the desirability of continuing to include the current selection of task types in future examination papers.

The term “sub-abilities” has been introduced to distinguish between the superordinate construct of the curriculum that was articulated in Chapter 4, and the underlying competencies that contribute different kinds of evidence of knowledge and mastery of English. They derive from the detailed learning content specified in CAPS and should not be confused with the traditional skills of reading, writing, listening and speaking.

In summary, the predominant purpose of the analysis in this chapter is to answer the following research questions:

- Using the articulated construct of a generic and differentiated language ability underlying CAPS and its predecessors as a basis,[32] what are the main sub-abilities that are being assessed in the English HL examination papers?
- Are the specifications of the sub-tests and task types adequate for providing evidence of the kind of language ability required by such a construct?
- Do the marking memoranda prevent rater bias and promote consistency of measurement?
- Is there adequate alignment between the language papers and the content and objectives of the new curriculum?

Satisfactory answers to the above are essential if the HL papers are to have any credibility and acceptability as high-stakes examination instruments (Bachman & Palmer 1996; Weir 2005; Weideman 2009a; Van Dyk 2010). Meeting the conditions for the responsible design of examination papers is also requisite for any comparison of standards across different language papers.

5.2 Methodology

A full selection of the November Grade 12 English HL Papers 1 and 3 covering the period 2008-2012 was analysed in detail. Each question of the respective papers was scrutinised from the perspective of a potential examinee, and possible answers compared to those elucidated in the marking memoranda. The purpose here was to evaluate the clarity of the questions, completeness of the memoranda, detect discrepancies between the question papers and memoranda, and express a view on the desirability of the set tasks and whether the papers were adequately aligned with the curriculum and its objectives. A limitation of the study was the exclusion of the Literature component of the curriculum that is examined annually in Paper 2. The reason for excluding Literature from the detailed analysis derives from the view that literary studies constitute a different construct and field of knowledge beyond the scope of this study.

[32] The construct has been articulated as "the assessment of a differentiated language ability in a number of discourse types involving typically different texts, and a generic ability incorporating task-based functional and formal aspects of language" (Du Plessis, Steyn & Weideman 2013: 20).

On the basis of the teaching plans and content of the CAPS curriculum (Department of Basic Education 2011a), six main categories were initially designated for the purpose of the analysis:

1. The reading process
2. Interpretation of graphic and visual texts
3. Vocabulary development and language use
4. Sentence structures and the organisation of texts
5. Process writing
6. Language structures and conventions during the writing process

Within each of these categories a numerical code was assigned to the detailed sub-abilities specified in CAPS under Subsection 3.2 (Reading and viewing), Subsection 3.3 (Writing and presenting) and Subsection 3.4 (Language structures and conventions). Subsection 3.1 (Listening and speaking) was excluded from the analysis as it forms part of school-based continuous assessment and is not assessed in Papers 1 and 3. It soon became apparent, however, that attempting to codify each of the examination items on the basis of one of the six categories above was a futile exercise. Just as listening, speaking, reading and writing take place in an integrated manner, many of the 89 sub-abilities listed in CAPS that were coded numerically cannot be compartmentalised to one of the six categories since they cut across more than one simultaneously. A decision was then made to move away from specifying the above categories separately and to allocate more than one numerical code per examination item. To do this the main purpose of each item was identified on the basis of the information provided in the task specifications and item prompts, as well as in the marking memoranda. The detailed categorisation and codes are provided in Appendix A.

In addition to identifying the dominant sub-abilities assessed per examination item, the number of items included in each section and paper per year of study was recorded, the number of marks allocated per item and whether marking was of a subjective or objective nature. Marking was designated as subjective where opinions were to be expressed or evaluated and on the basis of the guidelines contained in the memoranda. In those cases where a definite correct answer had to be provided without evaluative judgment on the part of the scorer, marking was considered to be of an objective nature. What should be kept in mind, however, is the fact that even where questions are not marked globally, the marker is given the right to consider other responses. There is a definite trend in the English HL examination to use the memoranda merely as a guide, and that the example

answers are by no means prescriptive or exhaustive. This naturally opens the door for subjective marking.

Further to the above, it was also noted whether an item required the construction or selection of an answer and items were categorised according to response type, i.e. what is referred to in testing terms as closed-ended or open-ended responses. The latter type requires of learners to construct their own answers, which may increase the possibility of subjective marking. Answers may vary in length from the writing of a phrase to the construction of a few sentences. Closed-ended questions, on the other hand, have higher reliability since there is a definite correct answer. Such items include those where examinees select an answer from options provided (e.g. multiple choice items), or where they have to select a word or phrase from a given text passage as the answer. Closed-ended questions are sometimes criticised for not providing an indication of productive or expressive ability (cf. ALTE 2005: 111-112). Usually a combination of different response types is used in language examinations such as the Grade 12 HL examination. The detailed classification of items is provided in Appendices B-D.

The results of the analysis of Paper 1 will be discussed first, after which attention will be devoted to the findings of the analysis of Paper 3. As mentioned, Paper 2 falls beyond the scope of the study and will not be included in this chapter.

5.3 Discussion of Paper 1: Language in context

Each of the three sections of Paper 1 will be discussed separately on the basis of dominant sub-abilities assessed, suitability of texts and visuals, the nature of marking, and other relevant aspects that could strengthen or undermine the validity and reliability of this part of the examination.

5.3.1 Section A – Question 1: Reading for comprehension

5.3.1.1 Dominant sub-abilities assessed

Text comprehension is the main task type in Section A. The sub-abilities of a generic nature tested in this section of Paper 1 during the five-year period under review are listed

in Table 5.1 according to frequency of occurrence (see Appendix A for the detailed classification system).

Table 5.1: Summary of dominant sub-abilities assessed in English HL Paper 1 Section A (2008-2012)

Code	Abridged sub-abilities	Frequency count
35	Express own opinion	30
46	Make sense of the text	17
26	Draw conclusions	9
83	Attend to word choice and language structures	8
19	Compare and contrast	6
23	Distinguish between denotation and connotation	7
27	Evaluate	5
37	Identify and explain use of figurative language	4
40	Infer	4
69	Summarise main and supporting ideas	4
17	Demonstrate an understanding of common phrases/idioms	3
43	Make connections	3
56	Read for main ideas	2
64	Retell a sentence using different words	2
8	Interpret/respond to language in cartoons/comic strips	1
7	Interpret/respond to images	1
10	Be aware of the socio-political and cultural background of texts and authors	1
25	Use direct and indirect speech	1
58	Recognise emotive and manipulative language/bias	1
73	Understand direct and implied meaning (critical language)	1
51	Use knowledge of prefixes/suffixes to determine meaning	1

Although 60 items were analysed, more than one sub-ability was tested simultaneously per item owing to the manner in which functionally defined lingual phenomena cohere. Obviously it would be impossible to cover the full syllabus in the scope of an examination and a selection of content has to be made. What is important is to include items from across the curriculum and not one specific area, so as to prevent unbeneficial washback and narrowing of the subject content. In this respect Hughes (2003: 54) advises against making an examination paper “highly predictable” and suggests including a wide range of sub-abilities, even those more difficult to test, and varying task types.

The analysis of the comprehension section reveals a representative selection of sub-abilities covered in the curriculum, but the content validity of the paper is weakened by the way questions and answers are formulated.

Examples of highly problematic items include the following three items from the November 2009 paper:

Item 1.1	Refer to paragraph 1. Explain why it is important that people cannot remember precisely when they played their first game. (2)
Memorandum:	We were too young to remember <u>when</u> we played the first game. OR Playing is an unconscious / instinctive / inherent activity. (2 marks for any one part)
Item 1.2	Refer to paragraph 2. Show how the writer proves his point that 'Play is essential to being human'. (3)
Memorandum:	Our earliest memories are of play. Sport is a form of play and a communal activity. It involves people. We start off by being involved in solitary sporting activities. This is later a shared activity but moves on to us being part of a team. (Award marks according to the depth of the candidate's response; mark globally) (3)
Item 1.3	Refer to paragraph 4. Explain why the writer feels it is less important to argue 'about who first discovered a game' than to experience the sheer enjoyment of it. (2)
Memorandum:	It is more important for people to enjoy being involved in sport than to be concerned about when it started. (2)

Item 1.1 is nonsensical, as is the awarding of 2 marks for "any one part". The next item illustrates the problematic nature of allowing global marking. In item 1.3, the proposed answer is merely a restatement of the item, just in different words. Again, it is unclear how the 2 marks will be awarded. Unfortunately, these three examples are typical of the kinds of items provided in the selection of English HL papers. Much more attention needs to be given to the clarity of questions and alignment of these with answers in the memoranda.

The following table lists specific items that were flagged as problematic in terms of their content, formulation or alignment with the memorandum.

Table 5.2: Summary of problematic items testing reading comprehension in English HL Paper 1 Section A (2008-2012)[33]

November 2009	
Item	**Problematic aspect**
1.1	Nonsensical item
1.3	Answer in memorandum repeats content of item prompt
1.5	Answer in memorandum repeats part of item prompt
1.11	Poorly formulated item and inappropriate answers in memorandum
November 2010	
Item	**Problematic aspect**
1.3	Poorly formulated text extract, item and answer in memorandum Mark is awarded for merely copying the sentence to be discussed
1.5.1	Item poorly formulated and does not correspond with answer in memorandum
4.1	Cultural bias requiring knowledge of Western marriage ceremonies; outdated image of cellular device
4.2	Cultural bias; error in memorandum (image does not indicate movement)
5.1	Incomplete answer in memorandum (both pronouns and their verbs need to be indicated)
5.7	Error in memorandum; more answers possible
November 2011	
Item	**Problematic aspect**
1.1.1	Misalignment with memorandum: Answer does not discuss religion metaphor
1.3	Memorandum contains irrelevant answer for 1 mark
1.5	Part of the memorandum incomprehensible
1.10	Memorandum incomplete: no mention of the Latin origin of the word
November 2012	
1.1	Poorly formulated: can be understood in different ways
1.3	Memorandum problematic: ignoring the ellipsis changes the meaning and no imagery can be discussed
1.4	Based on poorly written section of text with mixed metaphors: memorandum incomplete
1.5	Last part of memorandum answer not related to question
1.8	Too broad an item for 3 marks
1.10	Irrelevant item with unlimited possible answers
1.11	Memorandum problematic: comparison does not have to be made

At least one third of the items in the comprehension section (21 items of the 60 analysed) were found to be problematic. This is too high a percentage for a high-stakes language examination and suggests that insufficient procedures are in place to ensure the careful

[33] These are in addition to problems related to poorly written texts, open-ended subjective items and lack of indication of how marks will be awarded.

wording of items and responses and also that the system of moderating papers is not effective.

Apart from the above points of criticism, one sub-ability is severely over-represented in the comprehension section. Just over half of the items (52%) involve the expression of an opinion, making this a characteristic feature of Section A (e.g. “Do you agree…”, “In your opinion….”, “Suggest why ….”). Such item types require potentially subjective marking that could have a detrimental effect on scoring validity even before one considers variation relating to inter-rater severity and leniency differences. No marks are allocated for credibility of opinion or the masterful way a view is expressed through language – the mere fact that the learner has an opinion is what is assessed. As a result this kind of an examination item provides little indication of ability. The fact that so many of the items analysed fall into this category suggests that the final mark obtained per examinee may be inflated and unreliable. The extent of subjectivity involved in the marking is discussed further below in the section on mark allocation.

5.3.1.2 Reading texts provided

The difficulty of reading texts and kinds of themes used in the papers that were analysed are indicated in Table 5.3.

Table 5.3: Readability statistics of texts in English HL Paper 1 Section A (2008-2012)

2008	**Text A: History in the making (Discourse field: Politics)**	
	Number of words	320
	Flesch Reading Ease	68.5
	Grade level	6.8
	Text B: Untitled (Discourse field: Politics)	
	Number of words	358
	Flesch Reading Ease	47.4
	Grade level	12.5
2009	**Text A: The games that bring us together (Discourse field: Social)**	
	Number of words	432
	Flesch Reading Ease	75.1
	Grade level	7.0
	Text B: Youth sport for a healthy nation (Discourse field: Social)	
	Number of words	131
	Flesch Reading Ease	49.9
	Grade level	10.7
2010	**Text A: Comic strips and cartoons (Discourse field: Academic)**	
	Number of words	344
	Flesch Reading Ease	47.9
	Grade level	11.3
	Text B: Nelson Mandela comic book launched (Discourse field: Politics)	
	Number of words	419
	Flesch Reading Ease	41.1
	Grade level	12.4
	Text C: Untitled cartoon (Discourse field: Social)	
	No readability statistics available (too little text)	
2011	**Text A: Untitled (Discourse field: Politics)**	
	Number of words	715
	Flesch Reading Ease	43.5
	Grade level	12.7
	Text B: Invictus film poster (Discourse field: Politics)	
	No readability statistics available (too little text)	
2012	**Text A: The arts celebrate and inspire our democracy (Discourse field: Politics)**	
	Number of words	833
	Flesch Reading Ease	50.8
	Grade level	12.0
	Text B: R150m Soweto theatre packs entertainment punch (advertisement) Discourse field: Social	
	No readability statistics available (too little text)	
Average Flesch Reading Ease		**53.03**
Average grade level		**10.68**

The readability of texts can be calculated through measures such as Flesch Reading Ease and Flesch-Kincaid Grade Level (available in Microsoft Word). They provide an indication of the degree of difficulty of a text based on the average length of sentences and number of syllables in each word (Stockmeyer 2009). All the reading passages in the 2008-2012 English HL papers are adapted versions of authentic texts covering actual topics, but not all texts are well written. On the positive side, it is encouraging to note the tendency since 2010 to select more challenging texts. Low grade-level passages such as those that featured in the 2008 and 2009 papers do not belong in a Grade 12 HL paper.

Politics and sport seem to have been popular themes in Section A. However, language testers generally consider these to be undesirable topics as they may be biased culturally or in terms of gender. Texts dealing with war, politics, serious diseases or religious beliefs are also discouraged (Jennings, Fox, Graves & Shohamy 1999; ALTE 2005). The selection of texts in the English papers reveals a lack of awareness on the part of the item designers of the properties that make texts suitable for assessing reading comprehension. The inclusion of so many texts with political themes could be considered an attempt by authorities to indoctrinate or influence learners to uphold a particular political dispensation, as was the case under National Party leadership where education was used to uphold apartheid ideology. Better reading texts to select would be those rich in factual information that cover politically and culturally more neutral themes, or at least less contentious ones.

The length of the reading comprehension texts is also problematic. The directive issued in Circular E2 by the Department of Basic Education (2012a) to replace the two short reading texts with one longer passage is to be welcomed. More cognitive processing is involved with longer texts and the greater number of items that can be set facilitates generalisation of ability to other domains requiring reading. The use of short and undemanding texts compromises theory-based validity according to Weir (2005: 74). One concern, however, is the requirement to include a visual text related to the theme of the long passage. This may not always be practical. What is more, there are other ways available of probing the ability to interpret information that is visually and graphically presented, and at the same time thematically related to other texts.

5.3.1.3 Mark allocation

The trend in this section is to set between 10 and 12 reading comprehension items and to allocate between two to four marks per item. This is problematic as little indication is provided to examinees of how marks will be awarded. In some instances the items suggest that an opinion may be provided as an answer, but the memorandum provides a set of answers based on the text. In such cases it is difficult to determine whether the marking could be considered objective (according to the answers indicated in the memoranda), or subjective (depending on whether the marker found the opinion or answer of the candidate to be acceptable). These questions have been categorised as "uncertain". The following extract from the 2012 English HL Paper 1 serves as an illustration:

Text passage:	This April, South Africans were able to reflect on the past 18 years since we took that giant step towards becoming a country that can boast one of the most democratic constitutions in the world. Theatre in South Africa has always been a dynamic forum that has given us the courage to grapple with the state of the nation. Our writers, stand-up comedians, satirists and community-based artists have used their remarkable talents to create and nurture a climate that has allowed us all to become active participants in our democracy.
Question 1.1:	Why is theatre considered 'a dynamic forum' (line 4)? (2 marks)
Memorandum:	Theatre is considered a 'dynamic forum' as it has nurtured a climate of democracy. Those involved in the theatre have encouraged us to become participants in this democracy. [if a candidate explains the concept of 'dynamic forum', award 2 marks.] [if a candidate lifts directly from the passage, do not award more than 1 mark.]

The problem with this question is that it is formulated in too general terms. Some learners may give their own opinions, while others may simply quote a phrase from the text. There may be a number of unanticipated responses different to those contained in the memorandum. The possibility exists that an acceptable answer may be scored as incorrect by markers who adhere strictly to the memorandum. It is also highly unlikely that different markers would allocate marks in the same way. Items which are open to more than one interpretation can undermine the reliability of the examination paper (see Hughes 2003: 46). The point to be made here is that there needs to be consistency in the way an item measures knowledge or ability. Without a measure of reliability it would be impossible to make any inferences of language ability and to compare the overall performance of candidates.

In the example question provided above, it is further important to note that the text passage is poorly written and the example question anticipates a connection between the new South Africa and theatre, without the text itself providing any coherent link. Moreover, the text expresses the opinion of the author that theatre has contributed towards democracy. This is not a fact, and well informed students would be aware that theatre in South Africa has not "always been" a platform of democracy and was used in some productions to uphold the apartheid political dispensation. It is also problematic that full marks can be allocated where candidates explain what the words "dynamic forum" mean (e.g. a lively platform) without placing this in the context of drama and theatre.

Other more specific formulations of question 1.1 would be:

> Explain in your own words what is meant by the phrase 'a dynamic forum' (1 mark), and why the author of the text considers theatre to be a 'dynamic forum' (1 mark).

Or

> Explain in your own words what is meant by the phrase 'a dynamic forum' (1 mark), and quote a section from the first paragraph to show how this applies to the arts (1 mark).

Or

> Why does the author of the text consider theatre to be important? Mention two reasons from the first paragraph (1 mark each). Use your own words and do not quote from the passage.

Where a question counts more than one mark, and for the purposes of ensuring equitability of scoring, learners need to receive an indication of how the marks will be allocated (as illustrated in the three alternative items above). There is also a considerable difference in ability between recognising or quoting a suitable phrase, and using original wording to construct an answer. Allowing quotes is acceptable to illustrate understanding to some extent. However, if the intention is to assess both receptive[34] reading comprehension ability and productive or expressive language ability through the construction of a motivated written response, the mere lifting of phrases or clauses from the text by matching words contained in the question with those contained in sections of the reading passage should not be allowed. In this instance the short answer questions would simultaneously serve as writing tasks.

[34] The term "receptive" is used to denote an ability that is not manifest directly in overt behaviour. For example, when a candidate engages in listening or reading activities, it is impossible to determine whether the person has understood without setting additional tasks that not only involve listening or reading, but result in forms of behaviour that demonstrate the successful use of these abilities. In the instance of "productive" ability, on the other hand, the evidence of competence is immediately observable in the stretches of writing or speech produced (cf. Hughes 2003: 136 ff. for a discussion).

The following table provides a summary of how marks were allocated in Section A for the five-year period 2008-2012.

Table 5.4: Summary of mark allocation for English HL Paper 1 Section A (2008-2012)

Mark allocation	Number of items counting 1 mark	1
	Number of items counting 2 marks	31
	Number of items counting 3 marks	25
	Number of items counting 4 marks	3
	% of items counting more than 1 mark	**98%**
Marking subjectivity	Number of items scored subjectively	32
	Number of items scored objectively	20
	Number of items of uncertain nature	8
	Total number of items analysed	60
	Number of closed-ended items	1
	Number of open-ended items	59
	% of items scored subjectively **% of marks allocated subjectively (104/150 marks)**	**53%** **69%**

Much too high a percentage of marks derives from scoring of a potentially subjective nature, which indicates that those responsible for the design of the papers are not familiar with the essential principles of validity and reliability in language assessment. Further to this, the lack of specification on how marks will be allocated per item is unfair to examinees and can also have a negative effect on the notion of measurement unit equivalence when making comparisons across different HL papers, as we shall see in Chapter 6. If a measure of equivalence is to be established between the different papers, marks would need to be allocated on the same basis in all of the HL papers and a similar number of items would need to be set per section.

A further problematic aspect of Section A of Paper 1 is the lack of closed-ended/discrete item questions. Only one of the 60 questions required a single word for an answer. The inclusion of multiple choice type questions and items requiring a definite word or short response answer is essential to increase the objectivity of marking and reliability of measurement.

5.3.2 Section B – Question 2: Summarising in your own words

5.3.2.1 Dominant sub-abilities assessed

This section typically only includes one task in which examinees are required to produce a written summary of a short text provided. Traditionally, a summary task is considered a writing item (cf. Weigle 2002; Hughes 2003; Weir 2005), although it also assesses reading and other skills. The main sub-abilities generally assessed are the ability to make sense of a passage, identify the main point and relevant supporting points, and condense this information in a coherently written paragraph. However, the summary writing item in Section B is clearly not aimed at assessing these. Initially, in the 2008 and 2009 papers, learners were required to write a summary in a "fluent paragraph of approximately 90 words" (2008 paper, p. 7; 2009 paper, p. 7). However, in the 2010-2012 papers, examinees are able to write in paragraph or point form and the marking memoranda stipulate that any points from the text that the marker considers to be relevant can be used for the summary. A contradictory stipulation in the 2011 and 2012 memoranda is that "sentences and/or sentence fragments must be coherent" (cf. English HL Memorandum Paper 1, 2011, p. 6 and 2012, p. 5), without cognisance of the fact that coherence cannot be established on the basis of fragmented sentences or phrases. This may undermine the validity and reliability of the summary task. As a result, this item lacks theory-based validity and content validity and should be redesigned or removed.

5.3.2.2 Reading texts used

The table that follows summarises the information obtained from Section B for the five-year period 2008-2012.

Table 5.5: Summary of content of English HL Paper 1 Section B (2008-2012)

Year	Topic	Word count	Readability		Sub-abilities	Marking
			Flesch reading ease	Flesch-Kincaid grade level		
2008	Books and reading	359	54.2	10.1	44,46,69,88	Subjective
2009	2010 Soccer World Cup	330	58.4	9.4	44,46,69,88	Subjective
2010	Children's rights and freedom of action	349	62.7	8.2	44,46,69,88	Subjective
2011	Power of positive thinking	347	57.5	8.7	44,46,69	Subjective
2012	The meaning of face	370	49.1	11.4	44,46,69	Subjective

Specifications 44 and 69 (see Appendix A) relate to the sub-abilities to make notes and summarise the main and supporting ideas respectively. Specification 88 requires of learners the ability to produce texts that are coherent using conjunctions and transitional words and phrases. The latter was not required in the 2011-2012 papers. The table shows that the selection of short texts is also problematic. Theory-based validity requires an item in a language exam to elicit the same language and cognitive processing involved in authentic social settings (Weir 2005: 18), an aspect also referred to as interactional authenticity (ALTE 2005: 14). It is highly unlikely that any student would be required to summarise an already short extract in authentic academic settings. More careful consideration should be given to the texts selected for summarising and the response format, and for devising an authentic context in which summary writing may potentially be found useful.

5.3.2.3 Mark allocation

Marking is of a global nature and from 2009 has become more lenient. A holistic marking rubric is provided in the 2008 memorandum, assessing aspects such as sequencing of events, relevance of information and coherence. Moreover, up to 7 marks are deducted for exceeding the word count and marks are also subtracted for language errors and lifting whole sentences from the text. In the 2009 paper the summary task is mistakenly indicated as two separate items (2.1 and 2.2), when only one fluent paragraph is required, which may be confusing for the examinees. At first marking appears to be as strict as that in the

2008 paper since the memorandum states that only full sentences are to be accepted and language errors are to be penalised. However, on close analysis the memorandum shows that it is possible to obtain a score of 8 out of 10 simply by copying seven sentences from the passage. Only 2 marks are deducted for using "6-7 whole sentences" as quotes. In the 2010 paper, candidates are instructed to write a fluent paragraph, but the memorandum also allows other formats to be assessed and only 1 mark to be deducted as a penalty for "not presenting the summary in the required format" (p. 7). Here too it is possible to lift whole sentences and obtain at least 7 out of 10 even with the penalty for using quotes. In the 2011 and 2012 papers, examinees are asked to use their own words to summarise the text, but can write "either a fluent paragraph OR in point form" (English HL Paper 1, November 2011 and 2012, p. 7). This signals a change in the sub-abilities to be assessed. Although the 2011 and 2012 papers instruct learners to use their own words, the memoranda contain no penalties for failing to do so. There are also no penalties for exceeding the required length. There is further no indication of how sentence fragments will be assessed for language errors. The conclusion is inevitable that the papers and their memoranda have drifted over time to become more lenient. The errors in both the papers and the memoranda, however, show that their moderation is suspect.

The following excerpt illustrates the messy kind of scoring applied to the summary task in the 2011 and 2012 memoranda. One mark is to be allocated for each of seven points mentioned and three marks for language:

> **Distribution of language marks:**
> - 1-3 points correct: award 1 mark
> - 4-5 points correct: award 2 marks
> - 6-7 points correct: award 3 marks
>
> (English HL Memorandum Paper 1, November 2011 and 2012, p.6)

Presumably what is meant by the above is that for every two sentences or fragments, 1 mark is allocated for language use if no grammatical or spelling errors are made, but this is not stated specifically. Mark allocation does not appear to cover the organisation of the text. Technically, it would thus be possible to obtain a score of 7 out of 10 (70%) for the summary just on the basis of reproducing seven sentences or phrases from the text, without displaying knowledge of cohesion or coherence, and then earn language marks for copying these correctly. Equally problematic is the possibility that learners may far exceed the required number of 90 words, and still obtain the full seven marks if they have

listed seven points in the first 95 words of their summary, as is evident in the following extract from the 2012 memorandum:

- **Word Count:**
  - Markers are required to verify the number of words used.
  - Do not deduct any marks if the candidate fails to indicate the number of words used or if the number of words used is indicated incorrectly.
  - If the word limit is exceeded, read up to a maximum of 5 words above the stipulated upper limit and ignore the rest of the summary.
  - Summaries that are short but contain all the required main points should not be penalised. (English HL Memorandum Paper 1, 2012, p.6)

Section B of the paper serves very little purpose. In its current format it is uncertain what sub-abilities it assesses. Another disturbing aspect is the misalignment between the instructions given to examinees in the examination papers and directions given to markers in the memoranda. The summary writing tasks analysed are unable to generate *a priori evidence* of validity and cannot be considered reliable indicators of the ability to summarise texts.

From the analysis of the summary writing section, it is evident that there is a lack of conceptual clarity on what summary writing entails. Yu (2013) finds a number of incongruities in the conceptualisations of summary writing in literature and the way it is operationalised and assessed in language tests. Further to this, student responses to and interpretations of summary writing are influenced by their previous experiences and assumptions. Yu concludes that it is not a uniform ability or unitary process, but a multidimensional and unique kind of writing that requires integrated language ability (reading, analysing, condensing and restructuring in writing), making it a genre of its own, with potential contextual specifications that can obtain in different types of discourse. The summary of the plot of a novel may be different, materially and typically, from the summary of a political viewpoint, or of an ethical argument. This illustrates the need for a new understanding of the task on the part of the HL examiners, markers and examinees, as well as HL teachers – something that is evidently lacking at present.

If its typicality and format is dependent on discourse type, the ability to produce a coherent summary is a skill that students in all subject fields need. It features as a task type in prominent international tests such as the Test of English as a Foreign Language (TOEFL) (ETS 2012), the International English Language Testing System (IELTS 2016), and the Pearson Test of English (PTE 2016), all tests that place a high premium on the

validity, reliability, integrity and fairness of their tests. Preferably, a summary type task that assesses integrated language ability should feature in the HL papers, but in a responsibly designed format.

5.3.3 Section C – Questions 3, 4 and 5: Language in context

The heading for this section of the language paper is problematic, because it creates the impression that language can be divorced from context. Communicative language is always contextualised, as any authoritative book on language will confirm (Bloor & Bloor 2013, Eggins 2005, Halliday 2007). The entire HL examination should be considered language in context.

The current structure of Section C is as follows:

Question 3:	Analysing advertising
Question 4:	Understanding other aspects of the media
Question 5:	Using language correctly

In the discussions on Sections A and B above, separate subheadings were provided for the comments on the selection of text passages. In the discussion of texts and images used in Section C, however, comments will be combined with those on the dominant sub-abilities assessed as the texts and images are unsuitable for generating readability statistics.

5.3.3.1 Dominant sub-abilities assessed

The first two questions of this section focus on the ability to read texts of a visual nature and the third on correct language use. The interest in visual literacy and the study of the meaning of visual signs derives largely from the work of Ferdinand de Saussure who noted the distinction to be made between the visual image as the signifier and the concept it represented as the signified (cf. Culler 1986: 8). The problem, as Berger (1999: 71) points out, is that the "relationship between a signifier and signified is arbitrary, and therefore always open to question". Any examination questions pertaining to visual elements such as graphics, fonts, frame sizes, selection of objects, etc. will therefore be open to any number of different interpretations, an aspect that can undermine scoring validity, as we shall see further down. Another problem is that the reading of visuals such

as cartoons may depend on cultural and extraneous knowledge, making it a potentially unfair construct in an examination of English language ability. Perhaps the strongest reason for not including visuals such as photographs and cartoons in a language paper relates to the irrelevance of analysing such images in real-life contexts. It is a known fact that we do not spend time analysing cartoons or pictures. We may read them and appreciate them, but that is not the same as analysing them. Many of the tasks in Section C of Paper 1 lack authenticity and content validity since they do not replicate language use in what Bachman and Palmer (1996, 2010) refer to as target language use (TLU) in actual social settings, and are contrary to the clear intentions of the curriculum (CAPS) to promote the development of language use in real life. The exceptions are those that place the emphasis on the assessment of the meaning and use of words as signs.

The dominant sub-abilities covered in the selection of language papers are summarised below in order of frequency of occurrence and on the basis of the classification scheme used in Appendix A.

Table 5.6: Dominant sub-abilities assessed in English HL Paper 1 Section C (2008-2012)

Code	Abridged sub-ability	Frequency count
7	Analyse, interpret, respond to images and font types	24
35	Express own point of view supported by values, beliefs and experiences	22
8	Analyse, interpret, respond to language in cartoons and advertisements	19
65	Sentence construction (length and complexity)	11
83	Word choice and language structures	11
21	Concord	6
55	Punctuation	6
37	Figurative language and rhetorical devices	4
46	Make sense of the text	4
48	Nouns	4
64	Retell a story or sentence using different words	4
23	Denotation and connotation	3
68	Spelling	3
6	Analyse and interpret graphs	2
18	Commonly confused words, homophones/homonyms, synonyms/ antonyms	2
25	Direct and indirect speech	2
51	Prefixes, suffixes, roots of words	2
54	Pronouns	2
81	Verb forms and auxiliaries, tense and mood	2
1	Abbreviations and acronyms	1
2	Active and passive voice	1
4	Adverbs	1
3	Adjectives	1
16	Collocations	1
17	Common phrases, proverbs and idiomatic language	1
27	Evaluate	1
40	Infer	1
58	Recognise emotive and manipulative language, bias	1
79	Use textual context and cues to determine the meaning of words	1
72	Transition words/conjunctions	1
76	Use appropriate words, phrases and expressions in writing	1
84	Word order	1
89	Write texts that display own voice with style and register in keeping with the purpose of the writing	1

The detailed analysis shows that of the 89 sub-abilities specified in the curriculum, at least a third (33) were assessed to some extent in Section C of the English HL Paper 1 over the five-year period under study. Similarly to the findings pertaining to Section A,

there is an over-representation of sub-ability 35 (express own point of view), with more than half of the items (22 of 42) requiring examinees to express an opinion. Once again this has the potential to undermine the reliability of assessment owing to the subjective and global nature of marking involved.

A variety of visual texts was employed, including advertisements, photographs, posters and film images. Only one graph featured in the selection of papers (November 2010), which is surprising as the reading of graphs forms part of academic literacy and is a task type that can be scored reliably (cf. research findings on tests such as TALL, ICELDA 2016). One positive trend discernible since 2010 is the inclusion of visual images accompanied by lengthier text sections. This is to be welcomed, as the study of symbols and signs devoid of words cannot be justified in a language-specific examination and belongs rather in the field of Communication Science. The desirability of including visual semiotics as construct in an examination of language ability remains debatable, although the tasks are aligned with the prescribed curriculum.

The following items listed in Table 5.7 were flagged as problematic in terms of their content, formulation or alignment with the memorandum.

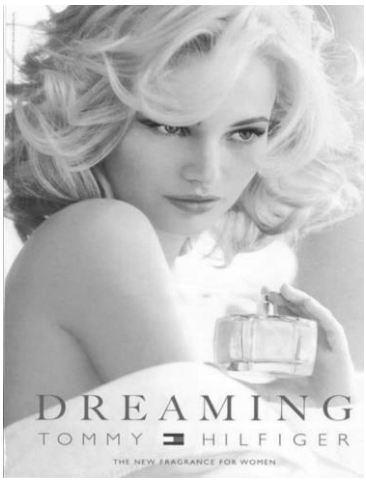
Table 5.7: Summary of problematic items in English HL Paper 1 Section C (2008-2012)[35]

November 2008	
Item	**Problematic aspect**
3.1.1	May be biased towards those with knowledge of Superman
3.1.2	May be biased towards those with knowledge of Superman
3.2.1	Gender and cultural bias: portrayal of women as sensual objects
3.2.2	Gender and cultural bias: portrayal of women as sensual objects
4.1.1	Item prompt provides answer
4.1.2	Memorandum incomplete: other answers possible
4.2.1	Poorly formulated (humour is in the book title) and memorandum incomplete: other answers possible
4.2.2	Memorandum incomplete (no mention of irony): other answers possible
4.2.3	Memorandum provides no example answers
5.1	No correct answer can be selected
November 2009	
Item	**Problematic aspect**
3.2	Gender and cultural bias: portrayal of men as sensual objects
5.9	All examinees awarded a mark even if they cannot provide the correct answer
November 2010	
Item	**Problematic aspect**
4.1	Cultural bias; outdated visual image
4.2	Cultural bias; outdated visual image
5.1	Incomplete answer in memorandum (both pronouns and their verbs need to be indicated)
5.7	Error in memorandum; more answers possible
November 2011	
Item	**Problematic aspect**
4.1.2	Memorandum only covers one of the two idiomatic expressions that need to be discussed
5.2	Error in memorandum
November 2012	
3.1	Visual image not clear enough to support item; memorandum does not provide for alternative answers
4.3	Memorandum incomplete: no mention of the fact that it may have been a bad dream and other answers are possible
5.3	Error in memorandum

In the 2008 paper, at least 10 of the 18 questions were problematic. However, because the HL papers are not standardised tests, each of the above items will not be discussed individually. A selection will be referred to in greater detail to illustrate why the items

[35] These are in addition to problems related to subjective scoring and mark allocation.

were flagged. The following examples from the November 2008 English HL paper (p. 11) contain content that may be biased towards certain examinees.



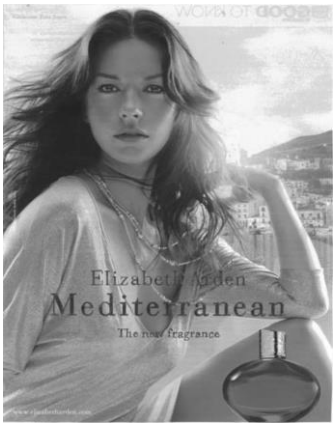
Text in advertisement:

DREAMING

TOMMY HILFIGER

THE NEW FRAGRANCE FOR WOMEN

Figure 5.1: Cultural bias in Text E, English HL Paper 1, November 2008, p. 11



Text in advertisement:

Elizabeth Arden
Mediterranean
The new fragrance

Figure 5.2: Cultural bias in Text F, English HL Paper 1, November 2008, p. 11

The following items were based on Figures 5.1 and 5.2:

Item: Discuss how the advertisers of the perfumes in Texts E and F appeal to their respective target markets, with reference to the following:
3.2.1 Words in the advertisements (2 marks)
3.2.2 Choice of models (2 marks)

Memorandum: 3.2.1 *Dreaming* – gives consumer idea of world of fantasy / bedroom.
New fragrance – something new on the market.
Name of famous designer – lends status / implies wealth.
Mediterranean – outdoor / exotic / historical.
Candidate may mention any of the above – one for each advertisement.
(1 mark per advertisement; credit discussion of *New Fragrance* if repeated for each.)

3.2.2 *Dreaming* – model appears romantic / coy / alluring / feminine / soft and gentle.
Mediterranean – model appears assertive / confident / forceful / forthright / challenging.
(Any description of beauty/sex appeal e.g. both models are beautiful/sexy, award full marks.)

These two items are not only gender insensitive, but culturally biased. To make the necessary associations, knowledge of the Mediterranean is required in the case of the second advertisement, as well as familiarity with the names of the designers. The suggested answers in the memorandum are all of a subjective nature. Would students who described the models as manipulative or dangerous receive a mark? Fortunately, all students can score full marks by simply copying the words "new fragrance" in 3.2.1. Tasks such as items 3.2.1 and 3.2.2 do not provide any evidence of language ability or critical language awareness and reveal a lack of conceptual clarity of constructs to be measured in this section.

Another example of a culturally biased item is the cartoon that forms part of item 4 of the November 2010 English HL paper (p. 10-11):

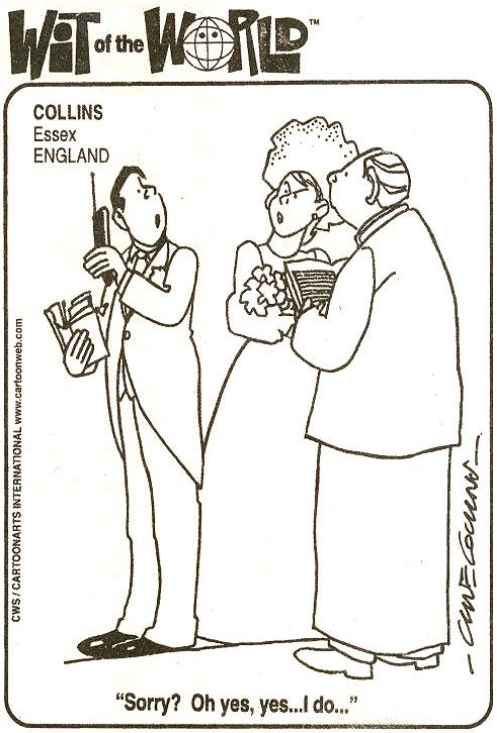


Figure 5.3: Cultural bias in Text G, English HL Paper 1, November 2010, p. 10

Examinees were required to do the following task based on Figure 5.3:

Item 4.1:	Discuss what the cartoonist is satirising (2 marks).
Memorandum:	The cartoonist satirises people's obsession with modern gadgets such as cell phones. This obsession interferes with traditional/normal considerations. People cannot be separated from their cell phones, even for something as important as their own wedding.

In order to discuss what the cartoonist is satirising, students would need to be familiar with Western wedding customs. It is also unclear what objects are being held in the hands of the gentleman on the left – a walky-talky or remote control device and manual? The memorandum states that a cellular phone is being held, but this is not clear from the outdated picture.

Apart from the problematic amount of subjective judgement involved in the assessment of the responses of examinees in questions 3 and 4, as well as the undesirability of including reading material with a cultural or gender bias, careful consideration needs to be given to the phrasing of items. Some items can be answered by repeating words contained in the prompt, for example item 4.1.1 of November 2008, p. 13:

Item 4.1.1	Explain the association between the illustration on the cover, which depicts a reading lamp, and the intention of the book. (2 marks)
Memorandum:	Novels that are a worthwhile read / illustrator has suggested reading by showing the lamp / light is associated with knowledge, and hence books and reading. (ANY ONE response)

Full marks can be allocated to learners for answering that the illustration suggests reading. The clue to the correct answer is thus provided in the phrase "reading lamp" contained in the wording of the item.

From the analysis it can be seen that items and suggested answers do not always provide for responses that reflect higher order thinking. Instead they tend to encourage lower order recall of information as they do not assess creative or critical language use. Item 4.2.2 of the 2008 paper serves as an illustration hereof.

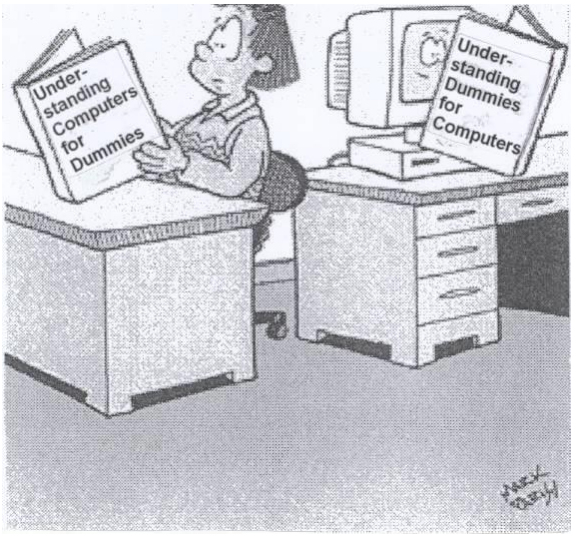


Figure 5.4: Illustration of lower order recall of information in English HL Paper 1, November 2008, p. 14

Examinees were required to carry out the following task based on Figure 5.4:

Item 4.2.2 Comment on the use of humour in the titles of the books. (2 marks)

Memorandum: Still necessary to consult books in order to operate computers. The computer, on the other hand, has to consult a book that explain *(language error in original memorandum)* how to deal with 'dummies' or people who are not familiar with computers.

Basically the answer in the memorandum simply repeats the titles of the books. There is no mention, for example, of the irony that humans invented computers, but need manuals to understand them, or that human beings themselves cannot be understood fully, even with a manual. The negative or at least ironical connotations associated with the word "dummies" are also not mentioned in the memorandum.

The decision to include so many visual texts in Section C must be questioned, especially since the item types focusing on slanting frames, fonts and facial expressions do not generate evidence of language ability. The study and analysis of visuals is motivated in CAPS by an erroneous assumption that "for many learners, the screen rather than the printed page is the source of most of their information" (Department of Basic Education 2011a: 23). This is a sweeping statement that fails to differentiate between different kinds of texts and purposes of reading. In the limited time that is available to assess *high* language ability, more fruitful task types than those illustrated in this chapter could be included.

The next section highlights the problematic way in which marks are awarded.

5.3.3.2 Mark allocation

One positive aspect in Section C is the increased number of items that can be marked objectively, as opposed to the low number included in Section A. However, even where marking is not of a subjective and global nature, no indication is provided to examinees as to how marks will be earned when answering the questions. The following table shows how the individual items in Section C contributed towards the final score obtained.

Table 5.8: Summary of mark allocation for English HL Paper 1 Section C (2008-2012)

Mark allocation	Number of items counting 1 mark	40
	Number of items counting 2 marks	37
	Number of items counting 3 marks	9
	Number of items counting 4 marks	2
	% of items counting more than 1 mark	**55%**
Marking subjectivity	Number of items scored subjectively: Question 3 Question 4 Question 5	 11 13 1
	Number of items scored objectively: Item 3 Item 4 Item 5	 9 10 43
	Number of items of uncertain nature	1
	Number of closed-ended items	35
	Number of open-ended items	53
	Total number of items analysed	88
	% of items scored subjectively **% of marks allocated subjectively (58/150 marks)**	**28%** **39%**

The higher amount of objective marking is attributable mainly to the items in question 5, which covers grammar. In both questions 3 and 4 there are more items requiring subjective marking than objective, making these two questions potentially unreliable. The weakened validity of scoring has to do with both the kinds of items and the way marks are awarded, as evident in the examples that follow.

Items such as 4.2.2 from the November 2011 paper have no incorrect answers. Markers thus have the prerogative to accept or reject the responses of the learners, or to give all learners full marks considering that the memorandum provides for global marking:

Figure 5.5: Global and potentially subjective scoring in item 4.2.2, English HL Paper 1, November 2011, p. 10

Examinees were required to execute the following task:

Item 4.2.2: The cartoonist does not show the mother-in-law in any of the frames. Do you think that this is an effective technique? Motivate your response. (2 marks)

Memorandum: Yes. The reader can supply his/her own idea of a hideous hat: this is more effective than drawing one./ The big gap in the relationship between Andy and the woman is suggested by her being out of the frames.
OR
No. I think it would have been very effective if the cartoonist had shown the mother-in-law wearing a hideous hat.
[Consider and credit other valid responses.]

From the memorandum we can see that the same marks are allocated for responses that reflect inferential and higher order thinking and those that merely require an opinion such as "No, the hat and mother-in-law should have been shown". The effect hereof is that the item does not discriminate between learners of differing ability and cannot be scored reliably. There is also a need to provide more comprehensive memoranda in order to assist markers of differing ability.

There is definitely a need for more detailed instructions in the papers and memoranda on how marks will be earned, because more than 1 mark is awarded for most of the items in questions 3 and 4. This will greatly assist both the learners and markers and increase the fairness of the assessment. The following example comes from item 3.2 of November 2009 (p. 11) and shows the discrepancies between item prompts and the suggested answers in the memoranda:

Item 3.2 Suggest why the advertiser asks the question **'Melting?'** in outstanding[36] letters. (2 marks)

Memorandum: By posing a question in bold/outstanding letters the advertiser is appealing directly to the reader. It draws our attention to the chocolate bar. The choice of the font creates the effect of a chocolate that is melting.
(Could be linked to the illustration).

The memorandum provides three aspects that can be mentioned, although other answers are also possible. However, the item counts 2 marks. Rephrasing the item to state "Suggest *two reasons* why the advertiser … " would clarify how much information needs to be provided. The same problem is evident in item 4.4 from the November 2012 paper:

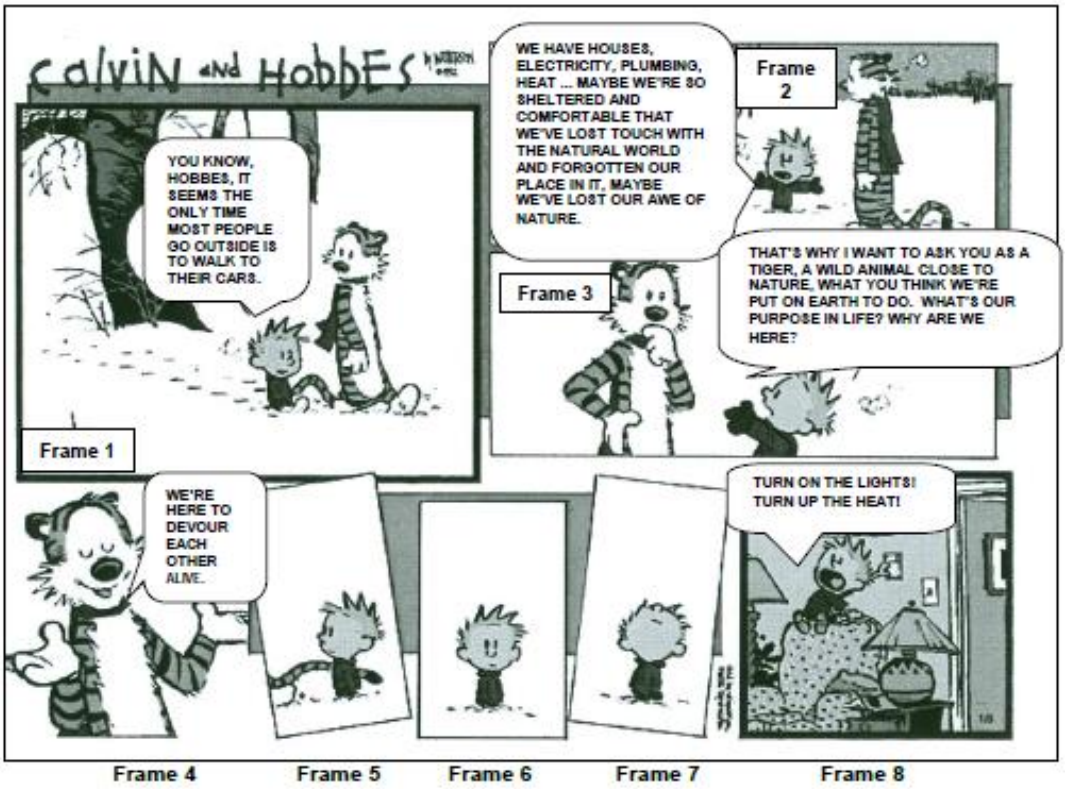


Figure 5.6: Item prompt and memorandum misalignment in item 4.4, English HL Paper 1, November 2012, p. 11

The detailed wording of the item prompt related to Figure 5.6 is provided below:

Item 4.4: Refer to frames 5, 6 and 7.
Comment on the effectiveness of the techniques used by the cartoonist in these frames. (3 marks)

Memorandum:

- The slanted panels/frames accurately capture Calvin's sense of isolation/inability to respond to the tiger's remark.
- The lack of text and speech bubbles is effective in conveying confusion.

[36] Presumably what is meant by "outstanding" is "boldface'.

- In frame 5, we only see a piece of the tiger's tail, which makes us wonder whether the tiger will act on what he suggests in frame 4.
- The direction in which Calvin is facing in each of the frames is an indication of his disorientation/fears/confusion about the friendship he shares with the tiger.
  [Award 3 marks for 2 techniques, well-discussed. Credit valid responses.]

Four suggested comments are provided in the memorandum, but the awarding of 3 marks for 2 techniques is problematic. It would be preferable to ask for a discussion of three effective techniques used by the cartoonist. Learners could then refer to any three of the following four aspects: the reason for the slanting frames; the absence of dialogue; the different directions in which Calvin gazes; or the disappearance of the tiger's tail.

The problems identified in the analysis of visual literacy items appear to apply to other HL papers as well. Moodley (2014: 204) found serious discrepancies in the visual literacy sections of at least six of the HL examinations as far as the "quality of questions for both technical knowledge and critical language awareness" was concerned as well as "cognitive demands". This supports the proposal of this study for a redesign of Paper 1. If this section of the examination is so problematic, it should be replaced by other language tasks that could potentially yield more useful information.

In contrast with questions 3 and 4, the section on correct language use in question 5 includes mainly 1-mark items. Very few discrepancies were identified here and the absence of subjective marking makes this section of the paper very reliable. Content validity is also high: the items are aligned well with the curriculum and examinees with a high language ability should be able to identify and correct errors in texts. One point of criticism is the failure to mention to examinees that they will be penalised for grammar and spelling errors in question 5. Once again, this points to a lack of understanding on the part of those responsible for designing the HL papers of important principles in language testing that help to ensure fair assessment practices.

5.3.4 Concluding remarks on Paper 1

The analysis of items reveals a number of recurring features in the analysed examination papers that impact negatively on the validity and reliability of Paper 1. Scores obtained by examinees cannot be generalised to non-examination domains for two main reasons. Firstly, foundational principles necessary to ensure the validity of content and alignment with constructs have not been applied, and secondly, the scoring of responses lacks

validity owing to insufficient task specifications and misalignment between item prompts and marking memoranda. In fact, the global nature of marking generally applied to Paper 1 creates an undesirable platform for mark inflation, the result of which can only be a set of highly unreliable examination papers with low credibility. The format of English HL Paper 1 should definitely be redesigned to include more relevant task types and response formats that generate evidence of high language ability.

5.4 Discussion of Paper 3: Writing

Writing assumes a prominent position in the HL curriculum and CAPS contains a plethora of genres and tasks aimed at helping learners develop dexterity in written communication. The premium placed on writing as a decisive component of language ability can be seen in the fact that Paper 3 contributes 25% of the total mark of the HL examination, as opposed to the 17.5% of Paper 1.[37] However, there are a number of material concerns related to the examination of writing in timed settings. In the first instance, CAPS advocates a process approach to teaching writing that cannot be replicated in an examination context in the current paper format, making it impossible to align Paper 3 with the curriculum. Secondly, owing to the separation of skills in the curriculum and teaching plan, very little classroom time may in fact be devoted to writing, making it potentially unfair to award so high a percentage of marks to writing. Even leaving aside the assessment of writing (e.g. summary writing) in other papers, it is disproportionately favoured by the allocation of marks for this paper. Thus, even before proceeding with the analysis, we have a further predicament with situational authenticity, a prerequisite for construct and content validity.

When including writing items in a language test or examination, every attempt should be made to ensure a conducive environment for writing:

> The context must be acceptable to the candidates and expert judges as a suitable milieu for assessing particular language abilities. The conditions under which tasks are normally performed should obtain as far as possible in a test of these abilities. A conscious effort should be made to build into tests as many real-life conditions as are feasible…….unless steps are taken to identify and incorporate such features it would seem imprudent to make statements about a candidate's ability to function in normal conditions in his or her future target situation. (Weir 2005: 56)

[37] The rest of the marks derive from Paper 2: Literature (20%), Paper 4: Oral (12.5%) and school-based assessment (25%).

To contribute towards situational authenticity, writing tasks should be representative of the kind of texts learners would be expected to produce in real-life situations. Ideally they should also be allowed to apply the same process approach to writing as in the classroom (or in real-life contexts) with sufficient time for reflection, consultation and revision. In view of the above, the selection of tasks in Paper 3 will be appraised mainly in terms of their authenticity and relevance, and whether they have been designed in a way that supports evidence generation of differential language ability. In the factual texts that examinees are required to produce, the opportunity should exist to display knowledge and application of the norms for language that are provided by and in the divergent lingual spheres articulated in CAPS, as opposed to just a generic ability.

Similarly to Paper 1, there are three examination sections. Section A covers creative writing in the form of essays, while Sections B and C assess longer and shorter transactional writing of a variety of text types. The system of codification of sub-abilities employed in Paper 1 will not be used for Paper 3, because productive language ability draws on multiple abilities simultaneously. Furthermore, since all scoring in this component is of a subjective nature, comments relating to scoring validity will be integrated with the general discussion.

5.4.1 Section A – Question 1: Creative writing

This section requires the writing of one 400-450 word essay. Six different types of essays are specified in the curriculum: narrative, descriptive, argumentative, discursive, reflective and literary essays (CAPS, Department of Basic Education 2011a: 28, 33-40). The last type forms part of Paper 2, leaving a selection of five other types for Section A of Paper 3.

The analysis revealed two major problems with Section A. Firstly, too great a choice of topics is provided – between eight and ten essay topics, some of which are based purely on visual images – which goes against the principle of limiting the choice of tasks to enable comparisons between the writing ability of candidates (see Hughes 2003: 94). Secondly, the construct exceeds language proficiency, and extends to areas such as imaginative ability and extemporaneous knowledge. Candidates are expected to write an

essay for 50 marks on esoteric topics such as the following extracted from the November 2012 paper (p. 3):

1.1 A path worth exploring
1.2 'When night falls over Africa, cities light up, creating patches of light visible from space. Compared to other places on the planet, the continent is pretty dark, but that is changing.'
1.3 To live is the rarest thing in the world. Most people exist – that is all.
1.4 If you run with the wolves, you will learn how to howl.
1.5 Tumbling from the heavens
1.6 *'The first wintry day*
You who sang like a robin
At last fell quiet.'
(Norman Morrissey)

Visual writing prompts are also provided as illustrated in the pictures below.

Figure 5.7: Visual writing prompts in English HL Paper 3, November 2012, p. 3

The following task specification for the visuals in Figure 5.7 was provided:

1.7 The pictures reproduced on pages 4 and 5 may evoke a reaction or feeling in you or stir your imagination.
Select ONE picture and write an essay in response.

It is obvious that the above tasks differ vastly in cognitive and communicative challenge and do not assess the same writing construct, making it impossible to compare performance of different examinees in an equitable manner. Highly intellectual and philosophical tasks (e.g. topic 3) and expository topics that require recall of factual or topical knowledge (e.g. topic 2) have different constructs and are much more demanding than a topic of a general nature where a personal experience or story can simply be shared. The varying levels of difficulty can introduce test bias (Bachman 2004: 156) and measurement error related to "construct-irrelevant variance" (Messick 1989: 34; Weir, Vidaković & Galaczi 2013: 439).

The analysis revealed three dominant kinds of essay writing in Section C.

Table 5.9: Types of verbal and visual essay writing prompts in English HL Paper 3 (2008-2012)

Year	General theme	Topical	Philosophical/Poetic
2008	Item 2		Items 1,3,4,5,6,7,8.1, 8.2
2009	Item 1		Items 2,3,4,5,6,7.1,7.2
2010	Items 1,5		Items 2,3,4,6,7.1,7.2
2011	Item 4		Items 1,2,3,5,6,7.1,7.2
2012		Item 2	Items 1,3,4,5,6,7.1,7.2

For the purpose of the above analysis, general topics were considered to be those which all students should be able to relate to, such as item 1 from the 2009 paper in which students could share their favourite days at school. Topical themes, however, required specific subject knowledge and recall of factual material, while philosophical and poetic topics demanded advanced and abstract reasoning and global imaginative ability. The visual prompts were difficult to categorise, as the learners could use them to sketch a personal narrative or produce a highly philosophical essay. Because of the vagueness of the prompts, they were placed in the third category. A limitation of the above analysis is the potential subjectivity of the researcher in categorising these items.

It should be borne in mind that examination tasks that require a global imagination and general knowledge, as well as those that depend on philosophical or poetic aptitude, could potentially disadvantage examinees and be considered unfair towards some learners. More general topics in no way preclude students from displaying poetic or philosophical prowess, but help to create a more equitable context for writing assessment amongst students with vastly different educational backgrounds and frameworks of exposure.

One striking commonality of all the items is the complete absence of writing specifications. Although the curriculum indicates that both generic and differential writing ability should be developed, it seems that the construct of writing has not been articulated well for the purposes of the examination and that there is hardly any assessment of differential ability. In fact, the choice of topic, style, genre, register, audience and purpose is left wholly to the students. The topics are broad and open to any

interpretation. As a result, the discourse fields are not defined and any register or style of writing could potentially be acceptable, but one cannot be certain. Through the inclusion of such vague writing prompts and the use of one generic marking rubric, the typical features of writing and normal conventions that apply are reduced to irrelevance.

The following brief comparison of writing prompts in a selection of HL examination papers in Afrikaans, English and Sotho shows how the construct may differ across language papers.

Table 5.10: Comparison of verbal and visual writing prompts in Afrikaans, English and Sotho HL Paper 3, Section A (November 2012)

	General theme	Topical	Philosophical/Poetic
Afrikaans	1.1, 1.5	1.4	1.2, 1.3, 1.6.1, 1.6.2, 1.6.3
English		1.2	1.1, 1.3, 1.4, 1.5, 1.6, 1.7.1, 1.7.2
Sotho	1.1, 1.3, 1.6	1.2, 1.4, 1.5	1.7, 1.8

The English and Afrikaans papers reveal a preference for philosophical and poetic writing ability as the main indicator of writing proficiency, whereas the Sotho paper provides a broader variety of writing prompts and the sharing of observations and experiences of a general nature referred to as "stream of consciousness" personal reflection (see Weigle 2002: 8), such as item 1.3:

1.3 It is your desire to see yourself being a star in one of the areas of entertainment. Write an essay by completing the following heading, and explain what you would like to be.
If only I could be a star...

Situational authenticity is highly problematic in Paper 3. Essay topics do not resemble the kind of writing ability required of students and post-matriculants, and as such Section A does not fulfil the notion of target language usage alignment, a condition for content and context validity (Weideman, 2009a). The configuration of predominantly narrative and philosophical types of topics in the English and Afrikaans papers carries us back to the England of the 1960s and the "personal growth version of literacy education" that privileged this kind of writing as definitive (Prinsloo 2004: 87). In her analysis of South African examination papers nearly a decade ago, Prinsloo (2004: 87) sharply criticises the kinds of essay topics set as being distanced from the concerns or likely interests of the

learners, echoing the essayist predilection for writing as reasoned social comment, a form of writing that assumes a middleclass location. On the matter of expecting expository essays that require topical knowledge, she describes this as doing little more than providing occasion for "expounding ignorance" (Prinsloo 2004: 87). From the analysis of writing topics it is evident that little progress has been made to decolonise essay writing in the HL examination. The topics cited in the content analysis have not only featured prominently in the examination papers spanning the 2008-2012 period, but continue to form a core part of current HL papers. The following topics come from the November 2015 English HL Paper 3:

1.1 There was no possibility of taking a walk that day.
1.2 The past is a foreign country.
1.3 'When she transformed into a butterfly, the caterpillars spoke not of her beauty, but of her weirdness. They wanted her to change back into what she always had been.'
'But she had wings.' (Dean Jackson)
1.4 Gold is the dust that blinds all eyes.
1.5 'There's a time for daring and there's a time for caution, and a wise man understands which is called for.' (In *Dead Poets Society*)

Apart from the elitist bias of the essay topics, the absence of clear task specifications is a serious deficiency – the only specification provided is the length of the essay (400-450 words). The liberty granted examinees to elect what kind of writing they wish to produce, makes it impossible to categorise any of the essay topics as narrative, descriptive, expository, discursive, reflective or argumentative, which detracts from the curriculum specification of the mastery of this much fuller range of genres. The writing prompts in the Sotho paper analysed attempt to achieve greater specificity by adding a qualifying sentence or two (cf. item 1.3), but still no indication is given of the purpose of the writing or audience, crucial aspects that would require differentiation of style and register, for example. If no purpose or audience is specified, then much leeway needs to be left for the learner's freedom to interpret the exact nature of the writing task, which consequently imposes an undesirable restriction on the examiner's responsibility to assess writing competence, if not in his or her ability to understand writer's interpretation.

The subjective nature of scoring essays is mitigated through the use of rating scales and marking rubrics that reflect the abilities to be assessed. This is a critical part of ensuring scoring validity (Weigle 2002: 109). The rubric used to assess the essay in Section A is provided on the following page.

Table 5.11: Analytical rubric used to score Section A, HL Paper 3, November 2012, p. 8 (Department of Basic Education 2015a)

	Code 7: Outstanding 80 – 100%	**Code 6: Meritorious 70 – 79%**	**Code 5: Substantial 60 – 69%**	**Code 4: Adequate 50 – 59%**	**Code 3: Moderate 40 – 49%**	**Code 2: Elementary 30 – 39%**	**Code 1: Not achieved 0 – 29%**
CONTENT & PLANNING **(30 MARKS)**	**24 – 30** -Content outstanding, highly original. -Ideas thought-provoking, mature. -Planning and/or drafting has produced a flawlessly presentable essay.	**21 – 23½** -Content meritorious, original. -Ideas imaginative, interesting. - Planning and/or drafting has produced a well-crafted and presentable essay.	**18 – 20½** -Content sound, reasonably coherent. -Ideas interesting, convincing. - Planning and/or drafting has produced a presentable and good essay.	**15 – 17½** -Content appropriate, adequately coherent. -Ideas interesting, adequately original. - Planning and/or drafting has produced a satisfactory, presentable essay.	**12 – 14½** -Content mediocre, ordinary. Gaps in coherence. -Ideas mostly relevant. Limited originality. - Planning and/or drafting has produced a moderately presentable and coherent essay.	**9 – 11½** -Content not always clear, lacks coherence. -Few ideas, often repetitive. -Inadequate for HL level despite planning/drafting. Essay not well presented.	**0 – 8½** -Content largely irrelevant. No coherence. -Ideas tedious, repetitive. -Inadequate planning/drafting. Poorly presented essay.
LANGUAGE, STYLE & EDITING **(15 MARKS)**	**12 – 15** -Critical awareness of impact of language. -Language, punctuation effectively used. -Uses highly appropriate figurative language. -Choice of words exceptional, mature. -Style, tone, register highly suited to topic. -Virtually error-free following proofreading and editing.	**10½ – 11½** -Critical awareness of impact of language. -Language, punctuation correct; able to use figurative language. -Choice of words varied and creative. -Style, tone, register appropriately suited to topic. -Largely error-free following proofreading, editing.	**9 – 10** -Critical awareness of language evident. -Language and punctuation mostly correct. -Choice of words suited to text. -Style, tone, register suited to topic. -Mostly error-free following proofreading, editing.	**7½ – 8½** -Some awareness of impact of language. -Language simplistic, punctuation adequate. -Choice of words adequate. -Style, tone, register generally consistent with topic requirements. -Still contains a few errors following proofreading, editing.	**6 – 7** -Limited critical language awareness. -Language mediocre, punctuation often inaccurately used. -Choice of words basic. -Style, tone register lacking in coherence. -Contains several errors following proofreading, editing.	**4½ – 5½** -Language and punctuation flawed. -Choice of words limited. -Style, tone, register inappropriate. -Error-ridden despite proofreading, editing.	**0 – 4** -Language and punctuation seriously flawed. -Choice of words inappropriate. -Style, tone, register flawed in all aspects. -Error-ridden and confused following proofreading, editing.
STRUCTURE (5 MARKS)	**4 – 5** -Coherent development of topic. Vivid, exceptional detail. -Sentences, paragraphs brilliantly constructed. -Length in accordance with requirements of topic.	**3½** -Logical development of details. Coherent. -Sentences, paragraphs logical, varied. -Length correct.	**3** -Several relevant details developed. -Sentences, paragraphs well constructed. -Length correct.	**2½** -Some points, necessary details developed. -Sentences, paragraphing might be faulty in places but essay still makes sense. -Length almost correct.	**2** -Most necessary points evident. -Sentences, paragraphs faulty but essay still makes sense. -Length – too long/short.	**1½** -Sometimes off topic but general line of thought can be followed. -Sentences, paragraphs constructed at an elementary level. -Length – too long/short.	**0 – 1** -Off topic. -Sentences, paragraphs muddled, inconsistent. Length – far too long/short.

The fact that the above rubric is used across all HL papers and for all types of writing may be problematic. Writing is also culture-specific and different writing conventions may apply to languages, a notion referred to as "contrastive rhetoric" (cf. Weigle 2002: 21). For example, English writing has been described as linear and hierarchically organised with a high premium on coherence, while Spanish writing may tolerate lengthy introductions and digressions to demonstrate topical knowledge (Weigle 2002: 21). The implication hereof is that rubrics would need to be designed per HL paper, and would also need to differentiate per rhetorical task. Argumentative writing, for example, requires a different rubric to narrative or descriptive writing.

The generic essay rubric does not support the validity of scoring and has not been empirically validated to show its reliability. A further problem is that even though there are three categories with seven levels of descriptors, there is still overlap of content. For example category three, "structure", overlaps with "planning" in category one, and relevance of information is assessed under both categories. It is also uncertain what is meant by "critical awareness of impact of language" as a descriptor in category two, "language, style and editing". The requirement to use figurative language in order to attain a high score in codes 6 and 7 can also be criticised as not being of relevance to all kinds of essays. A further note of criticism is that the distinctions between levels of writing are too fine, e.g. "choice of words suited to text" (code 5) and "choice of words adequate" (code 4).

On the positive side, the analytical rubric used to score the essay question provides more than one score per writing task. In theory this can contribute towards the reliability of the scoring, but then the descriptors that distinguish levels need to be made more pertinent. It would be preferable to reduce the number of levels to provide sharper distinctions. A downside of analytical scoring is that it is more time consuming than holistic marking. In addition, for the final mark to be considered reliable, more than one marker is needed per task. The reliability coefficient can be as low as 0.25 for a composition that is only scored once (Hughes 2003: 95). Although Umalusi does require the employment of a system of moderation to encourage more equitable scoring, where large numbers of candidates are involved the moderation of a selection of examination papers cannot compensate for rater bias and scoring inconsistencies. Careful consideration therefore needs to be given to the weighting of the third paper in view of the unreliable nature of scoring and the difficulty

of generalising writing ability assessed under timed conditions to a broader domain of writing.

Notwithstanding the above criticism of the essay writing component of the examination, creative writing ability is worth developing and should not be removed from the curriculum. It can be a rewarding experience when conducted under different circumstances. Learners should have access to a variety of resources and inspirational aids and be able to research their topics as most writers do. Within the constraints of a pressurised examination setting, however, other kinds of writing tasks are preferable.

5.4.2 Sections B and C – Questions 2 and 3: Longer and shorter transactional texts

Section B provides four choices of topics and requires texts of 180-200 words to be written, while Section C gives three choices and requires a text of 100-120 words. The different genres of transactional writing specified in CAPS (Department of Basic Education 2011a: 28, 34-39,) that provide the basis for assessing writing proficiency in this part of Paper 3 and their frequency of occurrence as writing tasks in the 2008-2012 examination papers are summarised below.

Table 5.12: Transactional genres of writing in CAPS and frequency of occurrence in Sections B and C of English HL Paper 3 (2008-2012)

Examination section	Genre	Frequency of occurrence
Section B	Formal letter	7
	Dialogue	2
	Informal speech	2
	Magazine article	2
	Review	2
	Formal speech	1
	Friendly/informal letter	1
	Interview	1
	Newspaper article	1
	Obituary	1
	Agenda of meeting	0
	Curriculum vitae (CV)	0
	Minutes of meeting	0
	Report (formal and informal)	0
Section C	Advertisement	4
	Diary entry	3
	Instructions	3
	Directions	2
	Flyer	1
	Postcard	1
	Poster	1

Sections B and C are supposed to assess transactional writing, which is defined simply as "functional writing" in CAPS (Department of Basic Education 2011a: 92). However, owing to a lack of conceptual clarity of constructs in this component of the examination, creative composition becomes the chief focus of many of the items. The following examples require imaginative rather than functional ability:

2.3 REVIEW

A CD and DVD store has just been opened in your area. You have been invited to attend its opening. As a columnist for the entertainment section of your local newspaper, write a review of the new store.
(November, 2010, English HL Paper 3, p. 6)

2.3 REVIEW

Recently, you dined at the Real Roots Restaurant, where cultural/traditional meals are served. A magazine/newspaper supplement has approached you to write a review of the restaurant. Write the review.
(November 2012, English HL Paper 3, p. 7)

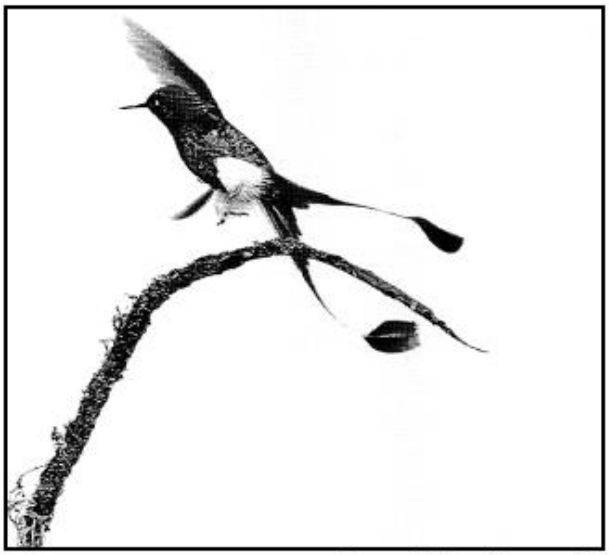
2.4 MAGAZINE ARTICLE

Young people of today experience an enormous amount of pressure to acquire material 'things'. Write an article for publication in your favourite magazine, in which you give advice on how to deal with this aspect of teenage life.
(November 2012, English HL Paper 3, p. 7

A visual prompt may also be used in this part of the paper as item 3.3 from the 2009 English HL Paper 3 illustrates.

3.3 ADVERTISEMENT

Use the photograph below to create a suitable advertisement for a product or service of your choice.



[Source: *National Geographic*]

NOTE: Your response must be limited to a written text. Marks will NOT be awarded for drawings, sketches, etc.

Figure 5.8: Visual prompt for transactional text in English HL Paper 3, November 2009, p. 9

A high level of creative language ability is needed to produce a text on the image of the bird and it would also be difficult to compare performance of examinees on such diverse items as the above examples. The only positive difference between the transactional section of the examination and the essay division is that at least an indication is given of purpose and audience in Section B, introducing an element of differential ability in the kind of writing required. Although formal letter writing seems to be the most common "transactional" task type, even in items such as the following there is an overlap of creative and functional writing:

2.2 PERSONAL LETTER

It is a long-established custom in your family to gather together for an annual festive function. You are holidaying with a group of close friends and the annual family event is approaching. You have been invited to extend your stay, but the time will clash with your family time. Write a letter breaking the news to your family. Ask for permission to stay with your friends.
(November 2008, English HL Paper 3, p. 8)

2.4 FORMAL LETTER

An end-of-the-year function at a restaurant developed into an embarrassing moment for you and your friends. Write a letter of apology to the manager explaining how you intend to remedy the situation.
(November 2008, English HL Paper 3, p. 8)

The following artificially contrived task types are devoid of all authenticity and depict little else than a lack of conceptual clarity of transactional writing as construct.

2.1 DIALOGUE

You have submitted your CV to a prospective employer and have been invited to a job interview. Write the job interview that takes place between you and the prospective employer. He/She has already read your CV and this is the meeting that follows.
NOTE: You are required to use the dialogue format.
(November 2009, English HL Paper 3, p. 6)

2.3 DIALOGUE

You are finalising your plans for next year, but your parent/guardian is not particularly happy about what you want to do. Write the dialogue that takes place between you.
(November 2011, English HL Paper 3, p. 7)

2.2 INTERVIEW

Two young children died and another was seriously injured after a main water supply pipe burst in a township in South Africa. As a newspaper reporter, you have been tasked to conduct an interview with the manager of the municipality. Write the interview.
(November 2012, English HL Paper 3, p. 7)

The inclusion of formal and informal speeches as task types is also questionable, since the writing and delivery of speeches already constitute part of the oral component of the curriculum and as such form part of school-based assessment, resulting in unnecessary duplication of assessment. It should be noted that the new curriculum states expressly that "these forms of writing are intimately connected with speaking, and should not be done purely as writing exercises" (Department of Basic Education, 2011a: 36). Yet there are examples in the HL papers of exactly this, such as item 2.4 from 2012 depicted in Figure 5.9.

2.4 SPEECH

Examine the photographs and text below.

[Source: Adapted from *Getaway*, April 2012]

Write a speech that you would present to RamsayMedia and in which you motivate your support for the most deserving cause.

NOTE: You may select only ONE of the photographs.

Figure 5.9: Visual prompts for transactional writing in English HL Paper 3, November 2012, p. 7

The examples of writing tasks provided in the preceding section are predominantly representative of Section B. If we look at the kind of writing required in Section C (100-120 words), we see that it is generally of such a basic and abridged nature (e.g. giving directions to reach a location, posting a notice or message on social media), that it can barely provide any indication of writing ability. This section is ill suited to the assessment of "high" language ability and should be removed. The following serve as illustrations of the rudimentary level of ability required in Section C:

3.1 FLYER

A clothing store is opening in your area. The manager has decided to hand out flyers to advertise the opening. Write the text for the flyer.
(November 2009, English HL Paper 3, p. 8)

3.2 POSTCARD

Write a postcard to your friend, telling him/her about your sports/cultural/educational tour in another province/country.
(November 2011, English HL Paper 3, p. 7)

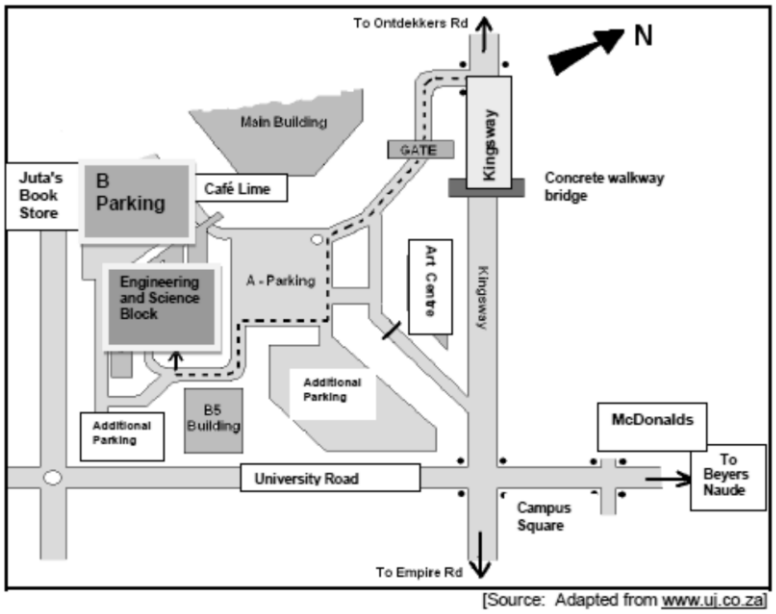
3.1 DIARY ENTRY

You have posted a message about a person, using social/digital/other media. Write a single diary entry in which you reflect on the message.[38] Note: Your tone may be informal but you may not use slang.
(November 2012, English HL Paper 3, p. 8)

Perhaps the most striking example of a writing task without any challenge is that of item 3.3 from the same 2012 paper.

3.3 DIRECTIONS

A friend is picking you up from the University of Johannesburg. He/she is travelling from Beyers Naude Drive towards Kingsway Road. After making a stop at McDonalds, he/she must pick you up in the B-Parking area on the campus.



[Source: Adapted from www.uj.co.za]

Using the map above, write down the directions that you will give him/her.

NOTE: Your route should include at least three turns and three landmarks. [20]

Figure 5.10: Example of rudimentary transactional writing in English HL Paper 3, November 2012, p. 9

[38] The writing prompt is ambiguous and seems to view the writing of a diary entry as synonymous with posting a message on social media. Matters are further confused by the instruction to "reflect on the message". Should examinees first fabricate a message to post, and then write a diary entry reflecting on that fabrication?

A brief analysis of writing tasks in the Afrikaans and Sotho HL papers of November 2012 reveals that the problems identified in the English papers are likely to be generalisable to the remaining HL papers. Compare the emphasis on creative composition in the following items from Section B (November 2012):

> 2.1 Write an article for the year book on an exceptional teacher who greatly influenced learners' lives. (Translated from Afrikaans paper, November 2012)
>
> 2.2 Your school has just returned from an educational tour, but unfortunately most learners have lost their belongings. Write a report in which you outline this incident. (Translated from Sotho paper, November 2012)

The following tables provide a summary of task types in the 2012 English, Sotho and Afrikaans HL papers.

Table 5.13: Analysis of writing tasks in Paper 3, Section B, November 2012 (Du Plessis & Weideman 2014: 140)

	Writing prompt	Formal letter	Interview	Review	Report	Formal speech	Informal speech	Obituary	Article	Full specifications	Authenticity
English	1	X								X	X
	2		X								
	3			X						X	
	4					X					
Sotho	1							X			
	2				X						
	3	X									X
	4			X							
Afrikaans	1								X	X	
	2	X								X	X
	3				X					X	
	4						X			X	

Table 5.14: Analysis of writing tasks in Paper 3, Section C, November 2012 (Du Plessis & Weideman 2014: 141)

	Writing prompt	Social media	Poster	Directions	Postcard	Diary entry	Invitation	Flyer	Full specifications	Authenticity
English	1	X								X
	2		X							X
	3			X					X	X
Sotho	1				X					
	2			X						X
	3							X		X
Afrikaans	1				X				X	
	2					X			X	
	3						X		X	X

The same lack of conceptual clarity of constructs is evident in the Afrikaans and Sotho papers, but at least the Afrikaans papers tend to provide fuller writing specifications. Only the formal letter writing tasks in Section B bear a close resemblance to the kind of longer transactional writing that learners may need to engage in after matriculating. Although the shorter transactional tasks required in Section C may have authenticity (with the

exception of the obsolete postcards and diary entries), as already pointed out they are too short to provide evidence of writing ability. This calls into question the content and context validity of Sections B and C.

On the matter of rubrics used for scoring Sections B and C, the same points of criticism levelled at the rubric used for Section A apply. There is overlap of content and insufficient distinction between levels. For example, in Table 5.15 how does one distinguish between codes 5, 4 and 3 on the basis of “fair”, “adequate” or “moderate” knowledge of text requirements?

Table 5.15: Analytical rubric used to score Section C, English HL Paper 3, November 2012, p. 10 (Department of Basic Education 2015a)

	Code 7: Outstanding 80–100%	Code 6: Meritorious 70–79%	Code 5: Substantial 60–69%	Code 4: Adequate 50–59%	Code 3: Moderate 40–49%	Code 2: Elementary 30–39%	Code 1: Not achieved 0–29%
CONTENT, PLANNING & FORMAT **(12 MARKS)**	**10–12** -Extensive specialised knowledge of requirements of text. -Exhibits a profound awareness of wider contexts in writing. -Disciplined writing – learner maintains rigorous focus, no digressions. -Total coherence in content and ideas, highly elaborated and all details support topic. -Evidence of planning and/or drafting has produced a flawlessly presentable text. -Has produced a highly appropriate format.	**8½–9½** -Very good knowledge of requirements of text. -Exhibits a broad awareness of wider contexts in writing. -Disciplined writing – learner maintains focus, no digressions. -Text is coherent in content and ideas, very well elaborated and all details support topic. -Evidence of planning and/or drafting has produced a well crafted and presentable text. -Has applied the necessary rules of format very well.	**7½–8** -Fair knowledge of requirements of text. -Exhibits a general awareness of wider contexts in writing tasks. -Writing – learner maintains focus, with minor digressions. -Text is mostly coherent in content and ideas, elaborated and most details support topic. -Evidence of planning and/or drafting has produced a presentable and very good text. -Has applied the necessary rules of format.	**6–7** -Adequate knowledge of requirements of text. -Exhibits some awareness of wider context in writing tasks Writing – learner digresses but does not impede overall meaning. -Text adequately coherent in content and ideas, some details support topic. -Evidence of planning and/or drafting has produced a satisfactorily presented text. -Has applied an adequate idea of requirements of format.	**5–5½** -Moderate knowledge of requirements of text. Response to writing task reveals a narrow focus. -Exhibits rather limited knowledge of wider contexts in writing tasks. -Writing – learner digresses, meaning vague in places. -Text moderately coherent in content and ideas, some details support topic. -Evidence of planning and/or drafting has produced a moderately presentable and coherent text. -Has a moderate idea of requirements of format – some critical oversights.	**4–4½** -Elementary knowledge of requirements of text. Response to writing task reveals a limited focus. -Exhibits a limited knowledge of wider contexts in writing tasks. -Writing – learner digresses, meaning obscure in places. -Text not always coherent in content and ideas, has few details which support topic. -Inadequate for HL level despite planning and/or drafting. Text not well presented. -Has vaguely applied necessary rules of format .	**0–3½** -No knowledge of requirements of text. -Exhibits no knowledge of wider contexts in writing tasks. -Writing – learner digresses, meaning obscure in places. -Text not coherent in content and ideas, has few details which support topic. -Inadequate planning/ drafting. Poorly presented text. -Has not applied necessary rules of format.
LANGUAGE, STYLE & EDITING **(8 MARKS)**	**6½–8** -Text grammatically accurate and brilliantly constructed. - Vocabulary is highly appropriate to purpose, audience and context. -Style, tone, register highly appropriate. -Text virtually error free following proofreading. -Length correct.	**6** -Text very well constructed and accurate. -Vocabulary very appropriate to purpose, audience and context. -Suitable style, tone and register considering demands of task. -Text largely error-free following proofreading and editing. -Length correct.	**5–5½** -Text well constructed and easy to read. -Vocabulary appropriate to purpose, audience and context. -Style, tone, register mostly appropriate. -Text mostly error-free following proofreading and editing. -Length correct.	**4–4½** -Text adequately constructed. Errors do not impede flow. -Vocabulary adequate for purpose, audience and context. -Style, tone, register fairly appropriate. -Text still contains few errors following proofreading and editing. -Length almost correct.	**3½** -Text is basically constructed. Several errors. -Vocabulary limited and not very suitable for purpose, audience and context. -Lapses in style, tone and register. -Text contains several errors following proofreading and editing. -Length – too long/short.	**2½–3** -Text is poorly constructed and difficult to follow. -Vocabulary requires some remediation and not suitable for purpose, audience and context. -Style, tone and register inappropriate. -Text error-ridden despite proofreading, editing. -Length – too long/short.	**0–2** -Text is poorly constructed and very difficult to follow. -Vocabulary requires serious remediation and not suitable for purpose. -Style, tone and register do not correspond with topic. -Text error-ridden and confused following proofreading, editing. -Length – far too long/short.

The same level descriptors are used to score Section B and C, with one difference: the amount of marks allocated and the addition of the following descriptor to Section C (Content, planning & format): "Exhibits a profound awareness of wider contexts in writing". What this refers to in transactional writing and how it will be assessed in such rudimentary tasks such as giving directions is not clear.

5.4.3 Concluding remarks on Paper 3

Owing to a lack of conceptual clarity of the construct of the HL examination and how a generic and differential ability should be assessed through writing tasks, the distinction between transactional and creative writing becomes obscured in Paper 3. The intention of the curriculum to have a multiplicity of genres, registers and discourse modes assessed is undermined. There is furthermore little difference between HL level and L2 or FAL kind of writing in the third examination paper. As pointed out in Chapter 4, at the level of a first language writing should be directed more towards the attainment of Cognitive Academic Language Proficiency (CALP) than Basic Interpersonal Communication Skills (BICS), and is associated with critical thinking and cognitive ability required for the expansion of knowledge. Weigle (2002: 5) explains that much of the focus in writing assessment at this higher level falls on the "originality of thought, the development of ideas, and the soundness of the writer's logic". This resonates with Bachman and Palmer's (1996) notion of interactiveness and the degree to which individual characteristics of an examinee are involved in executing a writing task. The emphasis in writing assessment at first language level does not fall explicitly on language knowledge, but other more strategic (metacognitive) aspects that indicate how the candidate is able to use the language he or she knows in a highly differentiated manner. However, the lack of writing specifications and the kinds of topics included in Paper 3 provide little opportunity to demonstrate such ability.

The kinds of tasks and topics that have been the feature of Paper 3 are decidedly unfair and require either poetic writing ability for the creation of literary artefacts far beyond the competence of most candidates, or the reliance on memory or fabrication of facts. This goes against the principles of natural language use and situational authenticity. With a view to establishing theory-based validity in writing tasks, Weigle (2002: 10) advocates not only determining the purpose of the writing, but the cognitive processing involved. She makes a

distinction between three levels of writing that feature prominently in most writing contexts, which may be summarised briefly as follows:

Table 5.16: Levels of writing tasks adapted from Weigle (2002: 10)

Level	Cognitive complexity	Purpose	Example task types
Level 1	Least demanding	Reproduction of information	Filling in forms Dictation exercises Giving directions
Level 2	Relatively demanding	Organising of information known to the writer	Compiling a report or summary Writing a letter
Level 3	Most demanding	Knowledge transformation and the generation of new ideas	Writing a persuasive academic essay Writing a critique

Creative composition does not feature in the above because it constitutes a distinct kind of imaginative construct that is unlikely to be expected of examinees in post-school domains. Consideration should be given to including only level 2 and 3 writing tasks in future examination papers and kinds of writing relevant to tertiary environments, a neglected area of academic development at school level (Bharuthram & McKenna 2012).

Scoring in Paper 3 is completely subjective, potentially unreliable and unfair. Examinees may elect to complete easier tasks for which they may potentially be awarded equally high marks as their fellow students who have attempted the more challenging tasks. Not only does this create an unfair basis for assessment, but it is problematic to infer that students who obtained high marks for Paper 3 will be able to produce appropriate writing in post-school contexts that require differential ability (e.g. at university). A further complication hereof is that institutions of higher learning are brought under the erroneous impression that their incumbent students are adequately prepared lingually speaking for academe, making it difficult to argue a case for systemic validity in the way HLs are taught and assessed.

Moreover, the weighting of Paper 3 needs reviewing. If we take into account the fact that we do much more reading, speaking and listening every day than writing, so many marks should not be allocated for writing ability. Chapters 6 and 7 debate this issue further and provide other alternatives. Writing is a multi-faceted construct simultaneously involving many of the sub-abilities already assessed in Paper 1, strengthening the case for an integrated approach to assessing language rather than devoting a separate paper to writing.

In fact, it can be argued that the process of arriving at producing a text in writing is so intertwined by prior processes of *finding* information (by listening, enquiring, discussing, reading, and so forth) and *processing* that information (again by digesting it, provisionally organising and analysing it, presenting it by articulating it, discussing and summarising it) that it would be difficult to separate it from other "skills" in the first instance (Weideman 2013). What is more, such separation can in fact impede rather than facilitate the instruction and development of writing, as well as its imaginative and adequate assessment.

As long as Paper 3 continues in its current format, it is likely to have a negative washback effect on writing development. It should also be clear by now that a curriculum such as CAPS cannot be expected to remedy the situation on its own. A more responsible approach would be to develop the required language skills holistically through a natural and integrated process of daily application, and that in all subject areas. In short, a return to the broader objectives of a mastery of a multiplicity of discourse and text types, genres and registers that are envisaged in the curriculum would go a long way towards ensuring a fairer assessment of language ability and increased proficiency in writing ability.

5.5 Conclusion

The fact that so many problems were identified in the analysis of English HL papers is particularly disturbing in light of the fact that the English papers are considered to be the flagship of the HL examination. The Umalusi report on the standards of the NSC HL examinations (Umalusi 2012a: 2) states that the evaluation of examination papers covering the period 2008-2011 revealed that the English papers were difficult in comparison to the other HL papers and contained "too few low level questions", but an "appropriate degree of challenge". Moreover, the English papers were the only language papers found to contain a sufficient number of grammar questions and adequate for preparing students for university study (Umalusi 2012a: 3). In view of the analyses of these papers in this chapter, these conclusions are somewhat surprising if not untrustworthy. The flawed nature of the Umalusi study was already pointed out in Chapter 1. In order to arrive at any conclusion as to the standard of the HL papers, a similar analysis to that undertaken in the current study would need to be undertaken of the remaining papers. As demonstrated through cursory references to writing tasks in other HL papers, the potential exists for the same problematic issues to

surface in a detailed study of the construct, content and scoring validity of the rest of the HL papers.

As long as the HL papers are unable to generate evidence of their validity and reliability as measurement instruments, any comparison of standards across languages would be fruitless. Further to this, even if such evidence were forthcoming, this would not be sufficient for comparing standards. Objective forms of measurement and statistical analysis would need to be introduced to provide a firmer basis for comparison. What is more, procedures will have to be introduced to compare these papers appropriately and adequately across years.

The next chapter examines how aspects other than similarity of construct may contribute towards measurement equivalence. Although all HLs share the same curriculum and all examination papers are required to comply with the format prescribed in CAPS (Department of Basic Education 2011a: 81), it is possible that the focus of the assessment may differ from one HL to another and that structural and technical elements could also make it difficult to determine whether papers are of a comparable standard.

6 Establishing a comparative structural basis for the HL papers

6.1 Introduction

The preceding chapters have focused specifically on the importance of achieving conceptual clarity on the underlying construct of the high-stakes Grade 12 HL examination and the need for the HL papers to assess similar abilities in view of the fact that they are premised on a common curriculum. Without similarity of construct and comparability of task specifications, no comparison of the results obtained in the HL subjects is possible. Chapter 6 pursues the matter of establishing a comparative basis for the HL papers further by investigating additional ways in which the HL examination papers can be placed on a more equitable footing. As a first step attention is devoted to the structure of the papers and task specifications. Examples from a selection of HL papers are provided first to illustrate how the papers may differ and then to identify factors that may obstruct technical forms of equivalence. Hereafter, the alignment of the prescribed format of the HL papers with syllabus weightings is investigated to provide a frame of reference for the possible restructuring of the papers. In particular, the amount of time devoted to the oral component of the curriculum is scrutinised. This is deemed necessary since oral assessment contributes 12.5% towards the final examination mark and the unreliability of this assessment could obstruct efforts to introduce greater equivalence of standard across all HL examinations.

6.2 Structural variations within a prescribed examination format

CAPS prescribes a standard format for all of the HL papers, which ensures that they have the same number of subsections and marks per section (Department of Basic Education 2011a) and that similar kinds of tasks are provided. Although some educators may be of the opinion that CAPS is too prescriptive in this regard, especially those in well-resourced upper- and middle-class schools who may prefer to show initiative and design their own teaching and testing materials, the comparability of format can be seen as an important step towards achieving greater equivalence of standard across all HL examinations. Chapter 2 discussed the disparate histories of the HLs and the lack of interest on the part of the education authorities with regard to the standard of teaching and assessment of the indigenous Bantu languages. The current practice to control the curriculum and examination from a central point can be seen as some indication of official interest in the standing of the

HLs, although it remains inadequate for the purposes of the further development of the languages as media of teaching and learning (cf. Du Plessis & Du Plessis 2015). Of importance in the current chapter is the extent to which there is uniformity of structural elements in the respective HL papers, and adequate task specifications.

The prescribed structure and task specifications for the three examination papers are provided in the section that follows. The relevant information has been captured in a similar tabular form to that used in CAPS, although information from more than one official document has been incorporated so as to reflect all existing guidelines in a single table.

Table 6.1: Structure and specifications of the NSC Grade 12 HL examination Paper 1 (Information obtained from Department of Basic Education 2011a: 81; 2012a)

SECTION	MARKS		TIME
A: Comprehension			
Question 1 Select TWO texts – ONE prose and ONE visual. The visual text must be related to the prose text. Length of Text A (Prose): 700-800 words (disjunctive orthography) 500-560 (conjunctive orthography) Length of Text B (Visual): Do not count the words in the visual. Learners should identify and explain the impact of techniques such as the use of font types and sizes, headings and captions, etc. There will be:- • comparative questions based on the two texts • a maximum of 2 open-ended questions • a maximum of 1 multiple-choice question	30		
B: Summary			
Question 2 The passage should not come from the comprehension text. Length of text (words): Approximately 350 Length of summary (words): 80-90 Candidates will be instructed to summarise in **point or paragraph** form. Provision for both answers will be accommodated in the memorandum, and candidates will not be penalised for either form of response.	10	70	2 hours
C: Language			
Questions 3-5 3 texts: 1 advertisement, 1 cartoon and 1 piece of prose Length of prose text: 250-300 words (disjunctive orthography) 120-150 (conjunctive orthography) Should assess vocabulary, language use, sentence structures and critical language awareness: Q3: Advertisement (combination of visual and written/verbal for 10 marks): Will test advertising techniques, language usage Note: There should be 1 open-ended question. Q4: Cartoon (10 marks): Will test visual and language usage. Note: There should be 1 open-ended question. Q5: Prose (10 marks): Will test language usage, grammar and editing skills in context. There will be a maximum of 2 multiple-choice questions in Section C.	30		

The examination specifications provided in CAPS do not incorporate amendments introduced through circulars issued by the Department of Basic Education since 2011. This

can be confusing to educators and examiners. Moreover, insufficient information is provided on how tasks are to be structured and some of the specifications in Table 6.1 are puzzling. For example, it is not clear why in Paper 1 multiple-choice questions are restricted to a maximum of one item in Section A and two in Section C. Well-designed multiple-choice items are used effectively in national and international tests to assess a range of abilities, including those related to text comprehension and text editing (ALTE 2005; ICELDA 2015). There also seems to be confusion on what constitutes an open-ended question. We see that a maximum of two open-ended questions are allowed in Section A and a minimum of two are required in Section C. However, as explained in the preceding chapter, multiple-choice and other items that require the selection of an answer from choices provided are considered closed-ended questions, whereas response formats that require the construction of an answer are open-ended items. If we examine Sections A and C of the November 2012 papers of three HLs, the mentioned specification of open-ended questions is not adhered to, and rightly so as it makes no sense:

Table 6.2: Item variations in Section A of Paper 1 in a selection of HL papers (November 2012)[39]

AFRIKAANS				
Item	Marks	Scoring	Type	Sub-ability
1.1	1	Obj	Open	46,73
1.2	2	Obj	Open	40,46
1.3	1	Obj	Open	56
1.4	2	Obj	Open	46
1.5	2	Obj	Open	11,35
1.6	1	Obj	Open	58,73
1.7	1	Obj	Open	17,37
1.8	1	Obj	Closed	17,37
1.9	1	Obj	Open	46
1.10	2	Obj	Open	36
1.11	1	Obj	Open	40,46
1.12	1	Obj	Open	46
1.13	1	Obj	Open	17,37
1.14	3	Sub	Open	35,37
1.15	1	Obj	Open	11,46
1.16	1	Obj	Open	46
1.17	3	Obj	Open	26,27,35
1.18	1	Obj	Open	17,43
1.19	1	Obj	Open	7
1.20	1	Sub	Open	7,35
1.21	2	Obj	Open	43

ENGLISH				
Item	Marks	Scoring	Type	Sub-ability
1.1	2	Sub	Open	46,35
1.2	3	Sub	Open	46,35
1.3	2	Obj	Open	37
1.4	2	Sub	Open	23,35,46
1.5	3	Sub	Open	35
1.6	3	Sub	Open	83,17
1.7	3	Sub	Open	19,35
1.8	3	Sub	Open	26,35
1.9	3	Sub	Open	83,35,37
1.10	2	Sub	Open	35
1.11	4	Sub	Open	19,35

SOTHO				
Item	Marks	Scoring	Type	Sub-ability
1.1	1	Obj	Open	11,46
1.2	1	Obj	Open	46
1.3	1	Obj	Open	46
1.4	1	Obj	Closed	17,37
1.5	1	Obj	Open	46
1.6	2	Sub	Open	35
1.7	2	Obj	Open	73
1.8	1	Obj	Open	11,40
1.9	2	Obj	Open	46
1.10	2	Sub	Open	35
1.11	2	Sub	Open	73
1.12	2	Sub	Open	35
1.13	2	Sub	Open	69
1.14	2	Sub	Open	19
1.15	1	Obj	Closed	17,37
1.16	2	Obj	Open	46
1.17	1	Obj	Open	46
1.18	2	Sub	Open	69
1.19	2	Sub	Open	35

[39] Sub-abilities in the fifth column refer to those categorised in Chapter 5 and listed in full in Appendix A.

Table 6.3: Item variations in Section C of Paper 1 in a selection of HL papers (November 2012)

AFRIKAANS				
Item	Marks	Scoring	Type	Sub-ability
3.1	1	Obj	Open	68
3.2	1	Obj	Open	1
3.3.1	1	Obj	Closed	73
3.3.2	1	Obj	Open	83
3.4	1	Obj	Closed	51
3.5	1	Obj	Open	3,83
3.6	1	Obj	Open	3,51,83
3.7	1	Obj	Open	18
3.8	1	Obj	Open	68
3.9	1	Obj	Open	18
3.10	1	Obj	Open	23
3.11	1	Obj	Open	81
3.12	1	Sub	Open	35
3.13	1	Sub	Open	37,73
3.14	1	Obj	Open	37
3.15	1	Obj	Open	65
3.16	1	Obj	Closed	73
3.17	1	Obj	Open	4,65
3.18	1	Obj	Closed	4,83
3.19	1	Obj	Closed	83
3.20	1	Obj	Open	25
3.21	1	Obj	Open	37
3.22	1	Obj	Open	81
3.23	1	Obj	Open	4,83
3.24	1	Sub	Open	35
3.25	1	Obj	Open	2,81,83
3.26	1	Obj	Open	55
3.27	1	Obj	Closed	23
3.28	1	Obj	Open	81,83
3.29	1	Sub	Open	7,35,43

ENGLISH				
Item	Marks	Scoring	Type	Sub-ability
3.1	2	Obj	Open	7
3.2	2	Sub	Open	8,35,83
3.3	2	Sub	Open	7,35
3.4	4	Sub	Open	8,27,83
4.1	2	Obj	Open	7,8,83
4.2	2	Sub	Open	7,35
4.3	3	Sub	Open	7,8,37,83
4.4	3	Sub	Open	7,35
5.1	1	Obj	Open	64,65,89
5.2	2	Obj	Open	25,65
5.3	1	Obj	Open	18
5.4	1	Obj	Open	21
5.5	1	Obj	Open	48
5.6	1	Obj	Closed	51
5.7	1	Obj	Open	21,65
5.8	2	Obj	Open	55

SOTHO				
Item	Marks	Scoring	Type	Sub-ability
3.1	1	Obj	Open	7
3.2	1	Obj	Closed	8,46
3.3	2	Sub	Open	8,46
3.4	1	Obj	Closed	46
3.5	2	Obj	Open	7
3.6	1	Obj	Closed	46
3.7	2	Sub	Open	35
4.1	1	Obj	Closed	46
4.2	1	Obj	Closed	46
4.3	2	Obj	Open	23,46
4.4	2	Obj	Open	35
4.5	2	Obj	Open	7
4.6	2	Sub	Open	35
5.1	2	Obj	Open	25
5.2	1	Obj	Open	81
5.3	2	Obj	Open	65,83
5.4	1	Obj	Closed	46
5.5	2	Obj	Open	65
5.6	2	Obj	Open	2

Differences in the number of items included in each of the three example papers analysed introduce a measure of inequality into the examination. It is obvious that students writing the Afrikaans HL paper in November, 2012, were required to provide considerably more responses than their peers in the English and Sotho streams. In total, the students writing the Afrikaans paper provided 51 responses, those writing the Sotho paper 38 and those writing the English 27 responses. The higher reliability of the scoring of the Afrikaans paper is also noticeable. In terms of measurement unit equivalence, the Sotho and Afrikaans papers that were analysed were more comparable with one another than with the English paper. The total number of closed-ended items in Sections A and C is almost the same (8 in the Sotho paper and 7 in the Afrikaans), and the mark allocation is limited to 1-2 marks per item, with the exception of one item (1.14) that contributed 3 marks in the Afrikaans paper. To make the papers structurally more comparable, it would be preferable for the specifications to be amended to provide a minimum and maximum number of items per section and to restrict the allocation of marks to 1-2 points per item for higher consistency of measurement.

Further to the above, the following three items from the November 2012 papers illustrate the necessity of wording examination items clearly so that examinees understand how marks will be allocated and how much information should be provided in the response. This is important for the purposes of establishing measurement unit equivalence across papers:

English HL Paper 1:
1.2 Explain how the artists become 'our nation's conscience'. (3 marks)

Afrikaans HL Paper:
1.17 Handwriting is an important key to expressing one's thoughts. Discuss the credibility of this statement within the context of the text passage. Motivate your answer with THREE references to information in the text. (3 marks)

Sotho HL Paper 1:
1.11 What kind of a person was Mmampitla when you look at her actions in general? Support your answer. (2 marks)

In general, the Afrikaans and Sotho items that were analysed in the November 2012 papers were worded much more clearly than the English items, and mark allocation corresponded with the instructions provided. Examinees may battle to dedicate an appropriate amount of time to complete each item in the absence of clearly formulated instructions that correspond with the marks allocated. Cognisance should also be taken of the fact that items which contribute more than 2 marks can amount to paragraph writing, which has the potential to change the construct and cognitive challenge of the task. If paragraph type comprehension responses are to be included in Paper 1, in addition to the summary writing task in Section B (which is uniform for all HL papers in terms of structure and measurement unit), this needs to be stated in the examination specifications (Table 6.1). Furthermore, the restriction on the number of open-ended items should be removed and a number of closed-ended items be allowed, irrespective of their format. The specifications should also contain an explanation of what is meant by open-ended and closed-ended items to assist educators and examinees, as well as those responsible for the design of the examination papers.

The following items from the November 2012 examination serve as examples of closed-ended items:

Afrikaans HL Paper 1:
1.8 Which word in paragraph 5 contrasts with the content of paragraph 6? (translated from original) (1 mark)

Sotho HL Paper 1:

1.15 Select the most appropriate answer from the following. Write only a letter (A–D) next to the answer that you choose.

To relax and be lax means…

A He should free horses to go and drink water.

B Sleep with feet up

C Take it slow and do nothing

D Jump into the water and swim

(translated from original) (1 mark)

A clear example of an open-ended item includes the following:

English HL Paper 1:

1.10 Discuss whether it would be more effective to include an image of the interior of the theatre or an image of the exterior of the theatre as a visual illustration in a newspaper article. (2 marks)

For the above item, the examinee cannot simply lift a phrase or clause from the text and is forced to construct an original answer. The following example, however, illustrates a poorly formulated open-ended item:

Sotho HL Paper 1:

1.2 What happened to Nyenye's girls when they left home? Mention just one incident from the text. (1 mark)

Not only can examinees simply lift a section of the reading text as their answer, the formulation is so general that at least eight answers are possible and all can be lifted from the text. The effect hereof is that what was meant to be an open-ended item becomes a closed-ended item with little challenge to the examinee.

With regard to the structural similarities of Section B of Paper 1, the analysis of the three selected HL papers and their marking memoranda shows that scoring is not applied consistently in this part of the examination either, causing further measurement unit dissimilarity.

Table 6.4: Structural analysis of Section B of Paper 1 in a selection of HL papers (November 2012)

	AFRIKAANS		ENGLISH		SOTHO	
Number of tasks	1		1		1	
Number of items per task	1		1		1	
Mark allocation per item	Identifying content	7 marks	Identifying content	7	Identifying content	7
	Language used	1-3 marks	Language used	1-3 marks	Language used	1-3 marks
	1-3 points with no errors	1 mark	1-3 points with no errors	1 mark	1-3 points with no errors	1 mark
	4-5 points with no errors	2 marks	4-5 points with no errors	2 marks	4-5 points with no errors	2 marks
	6-7 points with no errors	3 marks	6-7 points with no errors	3 marks	6-7 points with no errors	3 marks
					7 quotations	0 for language
					6 quotations and own words	1 mark for language
					4-5 quotations and own words	2 marks for language
					1-3 quotations and own words	3 marks for language
Response format	Open-ended		Open-ended		Open-ended	
Marks per task	10		10		10	

In theory, the specifications provided in the Sotho paper make it possible for examinees to obtain 7/10 marks for the content of the summary writing task, even if they simply lift seven points from the original text passage without using any of their own words. The fact that the English and Afrikaans memoranda, on the other hand, make no provision for quotations or points lifted from the text passage suggests that insufficient attention has been paid to providing full specifications for items, both in the examination papers and memoranda, as well as in the CAPS document. There is definitely a need to standardise specifications for the respective task types and to provide fuller instructions.

Although the current study does not cover the construct of the literature component of the curriculum, the format of Paper 2 and specifications for the design of items are provided

below for the purposes of completeness and the broader argument to achieve greater equivalence of standard and construct.

Table 6.5: Structure and specifications of the NSC Grade 12 HL examination Paper 2 (Information obtained from Department of Basic Education 2011a: 81; 2012a)

SECTION	MARKS		TIME
A: Poetry			
Questions 1-4 4 seen poems Two of the four questions must be answered (Marks: 2 x 10 = 20) Question 1: Essay (10 marks) (250-300 words disjunctive; 190-240 words conjunctive[40]) Questions 2-4: Contextual type questions (10 marks per question) **Questions 5-6** 1 unseen poem Either question 5 or 6 must be answered (Marks: 1 x 10 = 10) Question 5: Essay (10 marks) (250-300 words disjunctive; 190-240 words conjunctive) Question 6: Contextual type questions (10 marks)	30		
B: Novel			
Questions 7-12 Extracts from 3 different prescribed novels: 1 extract of 300-400 words (disjunctive) or 150-300 words (conjunctive) 2 extracts of maximum 200 words (disjunctive) or 150 words (conjunctive) 1 question must be answered in Section B There is a choice between an essay and a contextual question for each novel (25 marks each) Length of essays: Disjunctive: 400-450 words Conjunctive: 340-390 words	25	80	2½ hours
C: Drama			
Questions 13-16 Extracts from 2 different prescribed dramas: 1 extract of 200-300 words (disjunctive) or 100-150 words (conjunctive) 2 extracts of maximum 150 words (disjunctive) or 75 words (conjunctive) 1 question must be answered in Section C. If an essay question is completed in Section B, a contextual question must be completed in Section C. If a contextual question is completed in Section B, an essay question must be completed in Section C. There is a choice between an essay and a contextual question for each prescribed drama (25 marks each). Length of essays: Disjunctive: 400-450 words Conjunctive: 340-390 words	25		

[40] For a discussion on disjunctive and conjunctive writing systems in the indigenous languages and how words are joined to other words grammatically in conjunctive orthographies see Taljard and Bosch 2006.

The specification of the number of extracts from novels to be included needs clarification. The Afrikaans, English and Sotho examination papers each provided two extracts per contextual question, whereas the above specifications of the Department of Basic Education indicate the selection of one lengthier extract and two shorter extracts. Further to this, the considerable number of choices in Paper 2 results in very lengthy examination papers. For example, the 2012 Afrikaans Paper 2 was 27 pages long and the Sotho Paper 2 was 23 pages. Consideration should be given to reducing the number of choices by alternating the prescribed works so that each year a different novel and drama are examined, rather than including three choices of novels and two choices of dramas every year and having to set questions on all of these. From the perspective of the regulative condition of frugality discussed in Chapter 3, having to print so many pages per examination paper is not economical and amounts to a waste of expenditure.

On the structural side, we see in Paper 2 the same preferences for the number and types of items reflected in the analysis of Paper 1. Once again examinees writing the Afrikaans paper were required to provide more responses and scoring was limited to predominantly 1 mark per item, which increases the potential for high reliability of marking in this language paper.

Table 6.6: Comparison of number of items per contextual question in Paper 2 in a selection of HLs (November 2012)

	Question 2	Question 3	Question 4	Question 6	Question 8	Question 10	Question 12	Question 14	Question 16	Average number of items
Afrikaans	8	8	8	10	18	18	18	16	16	13.3
English	4	4	4	4	8	8	8	9	9	6.4
Sotho	6	5	6	5	14	14	15	14	15	10.4

Paper 2 includes almost exclusively open-ended questions with a high possibility for subjectivity on the scoring side. The English memorandum in fact makes provision for global marking of responses for the contextual questions, which resembles the marking of comprehension questions in the case of English HL Paper 1. The negative implications hereof for the reliability of the results obtained in the overall examination are obvious.

Lastly, the specifications for Paper 3 (Writing) are provided.

Table 6.7: Structure and specifications of the NSC Grade 12 HL examination Paper 3 (Information obtained from Department of Basic Education 2011a: 82; 2012a)

SECTION	MARKS		TIME
A: Essay			
Questions 1 8 topics, of which a minimum of 2 and a maximum of 3 will be visual stimuli Candidates will be expected to answer one essay question. Genres: narrative, descriptive, reflective, argumentative, discursive and expository. Length of essays: Disjunctive: 400-450 words Conjunctive: 340-390 words Assessment: Content and planning: 60% Language, style and editing: 30% Structure: 10%	50		
B: Longer transactional text			
Question 2 Four topics will be set from the categories indicated below. Visuals may only be used as supportive material. Category 1: Letter (informal or formal letter, business or friendly letter) Category 2: Speech or obituary Category 3: Written interview (formal) or dialogue (informal) Category 4: Report (formal or informal) Category 5: Business (memo, minutes or agenda, CV, formal letter of application) Category 6: Media (editorial, brochure, newspaper article, magazine article or review – formal) *(CAPS states two texts are to be produced. Circular E2 of 2012 states one text must be written):* Length of text: Disjunctive: 180-200 words Conjunctive: 100-120 words Assessment: Content, planning and format 60% Language, style and editing 40%	50	100	2½ hours
C: Shorter transactional/referential/informational text ***(Missing from CAPS, but included in Circular E2)***			
Question 3 Three topics will be set from the categories indicated below. Visuals may only be used as supportive material. Category 1: Advertisement, poster, flyer or invitation card Category 2: Diary or postcard Category 3: Instructions and directions One text to be produced. Length of text: Disjunctive: 100-120 words Conjunctive: 80-100 words			

One problem that emerged with the analysis of item specifications was that the information outlined in CAPS (Department of Basic Education 2011a: 81-82) did not correspond with that published by the Department of Basic Education in Circular E2 of 2012. The problematic aspects are indicated in italics in Table 6.7. Since both documents are official records, an attempt has been made to combine the information in the tables provided above. In CAPS there is only a Section A and B (p. 82), but Circular E2 of 2012 provides for a Section C as well (shorter text: transactional/referential/informational). CAPS provides for two transactional texts of 25 marks each in Section B, but Circular E2 of 2012 provides for one transactional task to be written in Section B (30 marks) and one in Section C (20 marks). A further problem is that in Section B the formal letter of application in category 5 overlaps with the tasks listed in category 1. Future versions of CAPS should incorporate all the amendments that have been introduced since the original publication of the curriculum and policy documents in 2011. If this does not happen, the authority of CAPS is undermined and assessment practice becomes policy.

The analysis of items in the English HL papers of November 2008-2012 and the Sotho and Afrikaans November 2012 papers shows that the specifications of Circular E2 of 2012 were adhered to up to 2013. However, in the November 2014 HL papers (Department of Basic Education 2015a) the format of Paper 3 changed and it seems that the specifications of CAPS are now being applied in Section B of Paper 3, while Section C has been removed. For the purposes of completeness of the current study, the format that was applied in 2012 in the three HL papers under discussion is provided below.

Table 6.8: Structural comparison of Paper 3 in a selection of HLs (November 2012)

	Afrikaans	English	Sotho
Section A Question 1	5 topics + 3 visual stimuli 1 task	6 topics + 2 visual stimuli 1 task	6 topics + 2 visual stimuli 1 task
Section B Question 2	4 topics 1 task	4 topics 1 task	4 topics 1 task
Section C Question 3	3 topics 1 task	3 topics 1 task	3 topics 1 task

The analysis of task and item specifications in Paper 3 of the three HL papers that were compared shows that examinees were required to answer the same number of questions and were also provided with the same number of choices of items. Scoring was carried out on a

comparable basis with each item contributing the same number of marks across all three language papers. Furthermore, identical rubrics were provided to score Paper 3, which may thus be considered to display structural and measurement unit equivalence.

From the information provided above, it can be concluded that on the whole the structure of the three examination papers is comparable across all HLs, with the exception of the variation in the number of items in Paper 1 and units of measurement applied in Papers 1 and 2. We can then state that the HL papers have a measure of structural similarity, but that this technical kind of equivalence has not been adequate to ensure similarity of construct and cognitive challenge, or reliability of measurement.

6.3 Alignment of teaching and assessment

Another aspect that has not been explored and that may have an impact on the ability to compare performance across languages is whether the syllabus weightings are reflected in the HL examination. This is an essential part of aligning teaching and assessment. To do this, a comparison is necessary of weightings in the examination papers in terms of hours of assessment and the designated classroom time (in theory) spent developing the "separate" skills of listening/speaking, writing and reading/viewing, as prescribed in CAPS (Department of Basic Education 2011a), as well as the time allocated for teaching and assessing the literature component. Further to this, we need to align weightings in terms of notional teaching and assessment hours with weightings in terms of mark contributions.

CAPS specifies that formal assessment is to take place during the year as part of school-based assessment, and at the end of Grade 12 during the exit-level external examination.

Table 6.9: Overview of formal assessment in Grade 12 (Department of Basic Education 2011a: 75)

Formal assessment		
During the year	End-of-year examination	
25%	75%	
School based assessment (SBA)	End-of-year exam papers	
25%	62.5%	12.5%
• 1 test • 7 tasks • 2 examinations (mid-year/test & trial/test)	**Written examinations** Paper 1 (2 hours) – Language in context Paper 2 (2½ hours) – Literature Paper 3 (2½ hours) – Writing	**Oral assessment tasks: Paper 4** Listening Speaking (prepared and unprepared speeches) The oral tasks undertaken during the course of the year constitute the end-of-year external assessment

The fact that oral assessment tasks, which collectively constitute Paper 4, contribute 12.5% towards the "end-of-year examination" is highly problematic. Unlike Papers 1-3 which are external assessments at the end of the Grade 12 school year, oral assessments take place throughout the duration of the teaching year and are administered internally. They can thus not be considered part of the external examination component, even if some system of external moderation is applied. If the results of Paper 4 are combined with those of SBA, as much as 37.5% of the final mark derives from marks awarded internally by teachers at the individual schools. This makes comparison of ability across the HLs an impossible feat – there are simply too many variables that play a role, especially in the light of the disparities referred to earlier in Chapter 2 that continue to undermine education at many schools. In fact, Umalusi admits that SBA is highly unreliable (Umalusi 2012a: 34).

Table 6.10 shows the nature of the continuous assessment tasks that contribute towards SBA and provides a summary of the teaching plan for the respective curriculum components for the year as set out in CAPS. Provision is made for 4.5 hours of formal contact time per week. Teachers may deviate from the schedule and devise their own teaching plans, but must ensure that the required assessment tasks are carried out. In this sense, the table should be seen as providing an indication of what the curriculum designers wish to prioritise, rather than as an indication of actual learning and assessment taking place in the classroom.

Table 6.10: Syllabus weightings in terms of teaching and assessment (Compiled from Department of Basic Education 2011a: 63-73)

CAPS curriculum components	**Notional teaching and assessment hours***	**Formal assessment tasks in term 1 (/marks)**	**Formal assessment tasks in term 2**	**Mid-year exam** at end of term 2**	**Formal assessment tasks in term 3****	**Formal assessment tasks in term 4**	**Exam preparation hours in term 4**	**External exam in term 4**
Listening activities	2	1 (/15)	0	0	0	0	0	0
Oral activities	13	1 (/10) prepared/ unprepared speech	1 (/10) prepared/ unprepared speech	0	1 (/15) prepared speech	Term 3 unprepared speech moderation	0	0
Reading and viewing (includes literature study, visual literacy, summary writing)	60	1 test: Language in context (/35)	1 Literature assessment comprising contextual questions (/10) and an essay (/25)	Paper 1: Language in context (/70) Paper 2: Literature (/80)	Trial exam**: Paper 1: Language in context (/70) Paper 2: Literature (/80)		Working through previous examination papers	Paper 1: Language in context (/70) Paper 2: Literature (/80)
Writing and presenting	60	2 tasks (essay /50) (transactional /25)		Paper 3: Writing (/100)	Trial exam**:: Paper 3: Writing (/100)		Working through previous examination papers	Paper 3: Writing (/100)
Total	135 hours (excluding exams) 21 hours (3 exam opportunities) Total: 156 hours	5 assessments 5 hours (min.) 135 marks	2 assessments 2 hours (min.) 45 marks	1 exam with 3 papers 7 hours 250 marks	4 assessments 2 hours for oral moderation 7 hours for trial exam 265 marks	1 assessment 2 hours for oral moderation 50 marks	16 hours	1 exam with 3 papers 7 hours 250 marks
Year mark	**25% (excluding oral work)**					**12.5%**		**62.5%**

*The HL teaching plan covers a total of 4.5 hours per week over a 30 week period (135 hours) and 10 additional weeks for taking examinations in all school subjects.

**Schools are required to write at least one internal examination in Grade 12 – at the end of the second or third term. If only one internal examination is administered, the second internal examination opportunity must be replaced by a test at the end of the relevant term (Department of Basic Education 2011a: 68)

From the above we can see that at least 48 hours of the 156 designated HL school hours over the scheduled 40-week period (i.e. 31%) are designated for assessment-related purposes, including internal and external examinations, formal tests, prescribed assessment tasks, moderation of orals and preparation for the external examination. All of these assessments are required to be recorded and moderated for the purposes of progression and certification (CAPS, Department of Basic Education 2011a: 74). If we add the additional hours needed to complete the required number of oral assessments per individual student, the actual percentage of time spent on assessment would be even higher. The point is that a minimum of one third of the total school time available is used for assessment purposes. Exactly how the oral assessments are supposed to be incorporated into the limited class contact time is not stated.

The oral component of the curriculum appears overly ambitious, attempting to cover in about 13 hours of contact time everything from teaching different features of discussions to delivering prepared and unprepared speeches and participating in dialogues, interviews, debates, panel discussions and even informal conversations. In addition, three different oral tasks per individual learner require formal assessment. A further complicating factor is that all of the oral tasks require different teaching strategies since the constructs vary vastly, especially where public speaking and speech writing are concerned. If we consider the impracticalities and logistical constraints of assessing large groups of students, in all probability teachers will resort to carrying out oral assessments without spending time on developing rhetorical ability, which would undermine the construct validity and reliability of the oral assessments. It may also be argued that at first language level there is no need to assess communicative oral ability in the FET phase (i.e. beyond Grade 9), unlike in the instance of second and foreign language learning. The oral and listening components could easily be integrated into other components of the HL curriculum where students could be given sufficient opportunities to exercise these skills during classroom discussions. It makes little sense to devote so many hours and marks on the assessment side to BICS-related oral tasks at first language level when the final outcome may not only be highly subjective and unreliable, but devoid of construct validity.

The impracticality of moderating the orals externally, as required by CAPS (p. 82), is obvious too and runs counter to the economics of language testing mentioned under the regulative conditions discussed in Chapter 3. Another reason to exclude orals is that there is already too much assessment of productive ability in the external examination, and too much potentially subjective scoring. In this regard, the removal of the oral component could reduce the

subjectivity and unreliability on the scoring side and provide a more equitable basis for comparing standards and performance across HLs. This important aspect will be attended to in the following chapter which covers possible alternative formats for the examination papers. It may be advisable to restrict all forms of SBA to a total of 25% rather than the current 37.5% because of the potential unreliability of scoring.

Although CAPS provides a clear indication of how the various aspects of the curriculum are to be covered during the course of a year, the teaching plan does not specify how to divide the available time between literature, reading comprehension, understanding visual texts and summary writing – all indicated in the curriculum under "Reading and viewing". Literature should be indicated separately as it constitutes a different subject area. This will help to determine whether all of the required HL components can in fact be covered in 4.5 hours per week, or whether additional time is needed. It is conceivable, for example, that teachers may devote almost all of the scheduled time to covering the prescribed poems, dramas and novels rather than attending to the language component of the curriculum if not enough time is available to cover all the prescribed parts of the language curriculum. If this is the case, Literature should rather be treated as a separate school subject and more notional hours should be designated for language teaching.

The following table provides an indication of the correlation between notional teaching and assessment hours and final mark contributions.

Table 6.11: Estimated alignment of teaching and assessment with mark contribution

Syllabus components	Notional teaching and internal assessment hours	External examination assessment hours	Mark contribution	
			SBA	External examination
Listening activities and orals	15	Cannot be determined	0	50 marks
Reading and viewing	60	Language: 3x2 = 6 hrs 28.5% of time	175 marks	70 marks
		Literature: 3x2.5 = 7.5 hrs 35.7% of time	195 marks	80 marks
Writing and presenting	60	Writing: 3x2.5 = 7.5 hrs 35.7% of time	275 marks	100 marks
Total	135	21	645 marks	300 marks
Final mark contributions			Reduced to 25%	Reduced to 75%

The curriculum component of "Writing and presenting" is accorded much prominence in the teaching plan, especially if we take into account the essay writing that forms part of the assessment of Literature and which is not reflected under "Writing and presenting". Almost twice as much teaching time is allocated for developing writing proficiency than for attending to general language ability or the study of prescribed literature. The notional teaching and assessment hours for the writing component are also disproportionately high if we consider the number of marks that writing tasks contribute towards the final score of Grade 12 learners. It is uncertain whether the scheduled teaching time will in actual fact be used for writing purposes, or whether it will be spent covering other aspects of the curriculum such as oral assessments and Literature. This is a further reason why language skills should not be separated in CAPS and why adopting an integrated approach to language learning is necessary. Integration of the so-called skills – or an acknowledgement that they can in fact not be de-integrated – would ensure that attention is devoted to all aspects and that on a regular basis.

6.4 Conclusion

The comparative analysis of the structure of a selection of 2012 HL papers in the present chapter reveals a strong preference on the part of HL examiners for direct testing of language ability through open-ended items. Despite this similarity, structural variations across the

Afrikaans, English and Sotho papers and a lack of item specification and measurement unit equivalence introduce a distinct element of unfairness in some parts of the HL examination.

This chapter has also examined the structure of the teaching plan provided in CAPS and pointed out the problematic nature of Paper 4, which covers school-based oral assessment of a highly unreliable and impractical nature. As long as the assessment of Paper 4 is highly unreliable, it will be impossible to compare standards and performance between individual examinees and across HLs. In addition hereto, little provision is made in the teaching plan for oral assessment, and by combining the teaching of the Literature component with language teaching in general, there is little certainty that all components of the curriculum will be covered. It would be preferable to adopt a teaching plan that integrates different language skills on a daily basis, and to provide a separate curriculum and teaching plan for the study of Literature.

In summary, two major findings of the study so far are that there is a lack of conceptual clarity on the constructs and sub-abilities to be measured in the respective examination papers, and that in their present format the English HL papers in particular are highly unreliable and unfair towards examinees. Task specifications are inadequate to guide examiners and educators, and guidelines for technical forms of equivalence need to be introduced. The possibility exists that similar findings would emerge in comparative validation studies of the remaining HL examination papers. The fact that the English HL examination papers are designed and administered under the auspices of Umalusi has not ensured adequate task specifications and alignment with the objectives of the curriculum. Consideration needs to be given to redesigning the HL teaching programme as well as the examination. The next chapter explores ways to attain greater comparability of standard and construct through a redesign of the examination papers that can improve the consistency of measurement in all HL papers.

7 Alternative formats for the HL examination papers

7.1 Introduction

This chapter examines the possibility of including in all of the HL papers a standardised component that can be subjected to statistical analysis. There are several reasons why this option should be considered. In the first instance, there is a need for more authentic forms of assessment to be used in the HL examination in order to argue a case for increased construct validity, as discussed in Chapter 5. Secondly, in the interests of attaining greater equivalence of standard there is a need to develop all of the HLs as academic languages, and at the same time to encourage the development of higher order thinking and inferential ability through specific task types. Lastly, the redesign of the papers is necessary in order to ensure similarity of construct and increased reliability of measurement across all HL papers. The introduction of a common standardised examination component could help to counter the subjective nature of some of the scoring and enable comparability of standard and construct in at least part of the Grade 12 HL examination. If sophisticated statistical modelling techniques are used, such a standardised component may even, if it is properly weighted in terms of its contribution to the total mark, be employed to equalise the marks in and across different HLs.

7.2 Increased authenticity of assessment

The increased emphasis placed on the socio-cognitive nature of language in later traditions of applied linguistics (referred to in Chapter 3), together with the need for increased task authenticity in language assessment (as discussed in Chapter 5), point to the desirability of changing the formats of the current HL papers. For the purpose of ensuring greater construct validity, more integrative forms of assessment of language ability that are in keeping with real-world language use contexts should be considered when endeavouring to measure overall ability (cf. Plakans 2012; Weir, Vidaković & Galaczi 2013). However, the teaching plan of CAPS (discussed in Section 6.3 of the preceding chapter) shows that in some parts of the curriculum there are still undesirable residues characteristic of first and second generation applied linguistics with their typical separation of skills. Overly ambitious direct oral assessments and a predilection for traditional essay writing in the form of creative composition – assessment types that hark back in many respects to 19th century traditions in language teaching – are characteristic of the speaking and writing components of CAPS. Moreover,

constant rehearsal (i.e. repetition) of skills for examination purposes appears to be encouraged throughout the school year with several rounds of examination-like assessments built into the teaching schedule. This is likely to narrow the curriculum and impede the development of a broad language ability. To counter such potentially negative washback from the HL examination, language skills should rather be integrated and examination papers should contain task types that draw on multiple skills simultaneously so as to reflect the way information is accessed and processed in natural language use settings.

As pointed out earlier, CAPS separates language skills into three main categories: "Listening and speaking", "Reading and viewing" and "Writing and presenting". In the language examination papers we see Paper 1 referred to as "Language in context", with a focus on reading comprehension and language structures (though writing is specifically also assessed), Paper 3 referred to as "Writing", with a focus on creative composition, and Paper 4 described as "Orals", which includes the assessment of "Speaking" and "Listening". If the current orthodoxy in language teaching is communicative and task-based, it would be more productive to adopt a skills-neutral approach both on the instructional and assessment side. Reading activities should preferably be combined with discussion and writing tasks. Attempting to teach writing separately runs counter to the fact that writing constitutes an important part of academic enquiry and is always preceded by obtaining information in various formats and engaging with subject content. Pot and Weideman (2015: 25) point out that curricula (such as CAPS) may separate language skills for organisational purposes, but that "in essence, these skills neither in actual fact exist as separate entities, nor are they easily distinguishable at a conceptual level"; this supports a more integrated approach to language assessment and language teaching. As others similarly have argued, "language development and agency do not emerge through writing in a structured university lecture hall or a tutorial alone. They are enacted and performed in all domains of the life of the individual" (Hibbert 2011: 33). It would make far more sense to integrate tasks that involve writing with other language related activities than attempting to coach certain types of writing on their own or to assess writing in a separate examination paper. Besides, the kind of writing that is the focus of much teaching in CAPS and that is assessed in the HL examination does not prepare students for academic writing (Du Plessis & Weideman 2014; Van Rooy & Coetzee-Van Rooy 2015). The continued emphasis placed on creative composition and economically-driven business communication of a most basic kind, as well as other types of (not always relevant) writing at first language level, may in fact be inhibiting the intellectualisation of the indigenous Bantu languages.

7.3 Developing the HLs as academic languages

A second motivation for an alternative format for the examination papers and the inclusion of a standardised component stems from the constitutional directive to maintain and advance all the HLs and ensure their equitable treatment as languages employed in high status spheres such as higher education. A common examination section that focuses on advanced language ability could strengthen the curriculum in achieving its objective to prepare students for higher education. In other words, by amending the format of the HL papers to include a focus on advanced language use and specifically the way that language comes to the fore in articulating the outcomes of inferential and analytical ability, there could be a positive washback effect on the teaching side. By using the HLs at a more advanced level, some of the stigma attached to the indigenous Bantu languages as supposedly (but erroneously) being unsuitable for advanced educational purposes could be eroded (cf. Alexander 2013a, 2013b; Webb 2013), and arguments based on prejudice that all the HLs are not suitable for developing "critical and close reading skills" (Umalusi 2012a: 7) be refuted. More advanced use of the HLs should facilitate their further development and use in the tertiary sector and other post-school milieus.

Another reason for developing all the HLs to a high level relates to the role they play in the acquisition of other languages and learning of subject matter in general. Although the HLs are taught as school subjects, only English and Afrikaans carry the status of academic languages and languages of learning and teaching (LOLT) beyond the intermediate phase (Grades 4-6). However, language acquisition research shows that a thorough knowledge of the first language (L1) or mother tongue (MT) assists with the scaffolding of learning other languages (Ellis 2008; Cummins 2011). This means that it would only be fair to prioritise all of the HLs for use as the LOLT so that all learners have an opportunity to benefit. Clearly, a concerted attempt should be made to develop all the HLs to a similar extent to that of English and Afrikaans, if any comparability of standard and construct were to be made.

The extent to which students are able to draw on the HLs as tools to acquire and learn other languages and engage with subject content (Cummins 2011) is in fact the objective of additive multilingualism. Nevertheless, as Balfour (2015: 188) points out, contrary to the Language-in-Education Policy, the South African school system encourages subtractive bilingualism. Owing to the emphasis placed on English FAL in the school curriculum, it seems that the position of most of the HLs has been weakened rather than strengthened. Note the subtle distinctions in

respect of the definition of the term HL and recent shift in policy emphasis (underlined for the sake of the discussion) in the following excerpts from CAPS and its predecessor, the 2008 version of the NCS:

Table 7.1: Policy shift in CAPS

NCS Learning Programme Guidelines (Department of Education 2008c)	CAPS (Department of Basic Education 2011a)
(No glossary or definition of HL provided –possibly considered self-explanatory and unnecessary)	**Glossary :** HL (see also additional language) – the language first acquired by children through immersion at home; the language in which we think (Department of Basic Education 2011a: 88)
HL: The learners' HL needs to be promoted, fostered and developed to provide a sound foundation for learning additional languages. <u>It may be used as the Language of Learning and Teaching</u>. Listening and speaking skills are developed and refined but the emphasis at this level is on developing the learners' reading and writing skills. (Department of Education 2008c: 8)	HL is the language first acquired by learners. However, many South African schools do not offer the HLs of some or all of the enrolled learners but rather have one or two languages offered at HL level. As a result, the labels HL and First Additional Language refer to the proficiency levels at which the language is offered and not the native (Home) or acquired (as in the additional languages) language. For the purposes of this policy, any reference to HL should be understood to refer to the level and not the language itself. (Department of Basic Education 2011a: 8)
The Department of Education's *Language in Education Policy* promotes additive multilingualism. This means that learners must learn additional languages while maintaining and developing their HL(s) at a high level. Additive multilingualism makes it possible for learners to transfer skills, such as reading, writing and speaking, from the language in which they are most proficient to their additional languages. (Department of Education 2008c: 7)	Additive multilingualism **–** when a person learns a language (or languages) in addition to his or her HL. This language does not replace the HL but is learned alongside it. In an additive multilingual programme, the HL is strengthened and affirmed while any further language learned is seen as adding value (e.g. all <u>Additional Languages, including the Language of Learning and Teaching, are taught alongside the HL</u> but do not replace it). (Department of Basic Education 2011a: 85)

The earlier NCS curriculum alludes to the fact that the HL is considered the language in which students are most proficient and that it may be used as the LOLT. Its successor CAPS, however, moves away from mentioning the HL as the LOLT: "all Additional Languages, including the Language of Learning and Teaching, are taught alongside the HL …" (Department of Basic Education 2011a: 85). However, notwithstanding the anomalies as to whether schools offer the languages first acquired by children through immersion at home as HL school subjects or whether they do not, and what the dominant language of a school-going child might in reality be, the fact remains that the LOLT that is used in the Foundation and Intermediate Phases of schooling should be a familiar language in which the young student is fully conversant, and this language should continue to be used as a medium of instruction at a high level beyond the early years of school education in order to be "strengthened and affirmed" (Department of

Basic Education 2011a: 85). This is not the case as a result of the insistence of the Department of Basic Education to switch to English FAL as the LOLT as early as possible (cf. CAPS for Foundation Phase, Department of Basic Education 2011b: 8), even after a mere three years of language learning. The effect of this is evident in the disparities reported in the annual results of the HL examination and differences in task complexity identified in the analysis of examination tasks. Although each HL is used as the LOLT in the Foundation Phase – and should be on a comparable footing at least at this early stage of language learning – it is clear that the development of a high level of language ability is not being encouraged across all HLs beyond Grade 3. As a result, all HLs have not developed a comprehensive vocabulary to a similar level as English and Afrikaans have and do not carry the status of academic languages. It would be preferable to rectify this inequality by developing and using all HLs to a far greater extent at school. There may also be much to gain from implementing a form of bilingual education in schools.

Cummins (2011: 166) points out the significant positive relationship that exists between developing academic skills in both the L1 (i.e. HL) and L2 (i.e. FAL) where there is sufficient exposure to these languages and motivation to learn them. Moreover, the educational success of additional language learners is positively related to continued instruction through the students' L1. The educational gains are thus increased when students receive instruction through their first languages for a longer period (Cummins 2011: 168). Although transition to a different LOLT that has been taught at an adequately high level alongside the initial LOLT is possible beyond the elementary phase of schooling, especially in middle-class and well-resourced settings, the question arises as to whether discontinuing the initial LOLT is counterproductive and serves to create more disparities. After all, the dominant or first language is supposed to be the language "in which we think" (Department of Basic Education 2011a: 88). In contradiction to this, CAPS advocates moving away from the HLs through its emphasis on "using the First Additional Language for the purposes of thinking and reasoning" (and that at an earlier stage than the previous curriculum) to enable learners "to develop their cognitive academic skills, which they need to study subjects like Science in English" (Department of Basic Education 2011a: 8). Unfortunately, social beliefs that English has more educational value than the remaining official languages undermine the objectives of the Language-in-Education Policy (McKay & Rubdy 2011) and lead to the neglect of the status and advanced use of the indigenous Bantu languages in particular. It is therefore not surprising to note the confession in CAPS that by Grade 10 "the reality is that many learners still cannot

communicate well in their Additional Language at this stage" (Department of Basic Education 2011a: 9), even though it has been taught and used as the medium of instruction for almost seven years by then. An inability to "communicate well" implies an inability to cope with cognitive academic demands in a language that has not been mastered at a high enough level. From this we can see that the quality of language education in both the HL and FAL at school is of paramount importance. It may thus be more fruitful to allow at least some school subjects to be taught through the HL beyond the Foundation Phase. This could encourage bilingualism and parity of esteem and support the development of a broad academic vocabulary in more than one language.

Ironically, the 2008 curriculum admits the disadvantageous position of the indigenous Bantu languages and the problematic aspect of not educating students through their HLs, but little effort has been made since to change this:

> **2.2.7 Human rights, inclusivity and environmental and social justice**
> In the past, languages in South Africa did not enjoy the same status. Learners and teachers need to be aware of the historical disadvantage African languages have suffered. Afrikaans and English were given special privileges both in terms of language teaching and their traditions and cultures. These inequalities are addressed by new national language policies. Schools should not further entrench the inequalities of the past by, for example, only teaching African languages as second or third additional languages and giving more time and attention to English and Afrikaans.
>
> A further problem in terms of language and human rights is that a large majority of learners in our country are not taught through the medium of their HL.
> (Department of Education 2008c:13)

Although in the new political dispensation the indigenous Bantu languages are not taught only as second or third languages, more time is still devoted to English and Afrikaans, and most learners continue not to be taught "through the medium of their HL" (Department of Education 2008c: 13). Despite the different intentions of the education policy, in other words, some of the inequities of the past dispensation continue. It is thus not surprising that higher education studies reveal that university students have low academic literacy and language proficiency levels (Bharuthram & McKenna 2012; Lewin & Mawoyo 2014; Rambiritch 2014), and that they battle with inferential ability and analytical thinking (Pretorius 2002; Grosser & Nel 2013). These are some of the unfortunate results of "assimilationist subtractive submersion education" that uses a global language such as English FAL as the main LOLT at the expense of the MT or L1 language of the student (Phillipson & Skutnabb-Kangas 2011: 34). Language policy scholars Phillipson and Skutnabb-Kangas (2011: 34) vehemently oppose this situation,

referring to it as "linguistic genocide" and a negation of basic linguistic human rights in education:

> Without binding educational linguistic human rights, especially a right to mainly mother tongue-medium education in state schools, with good teaching of a dominant language as a second language, given by competent bilingual teachers, most indigenous peoples and minorities have to accept *subtractive* education through the medium of a dominant/majority language. They learn a dominant language at the cost of the mother tongue(s). These are displaced, and later often replaced by the dominant language. (Phillipson & Skutnabb-Kangas 2011: 34)

The result hereof is that insufficient attention is paid to the importance and sustained development of the MT or HL and the bi-/multilingual potential of students. Another crucial factor pointed out by Phillipson and Skutnabb-Kangas, which has been overlooked in the South African context, includes the necessary competence and skill of "bilingual teachers". The extent to which the HL educators are adequately skilled to teach the HLs and operate in bilingual/multilingual classrooms is an area that warrants separate and urgent scrutiny. If educators were to receive far more intensive training in respect of both teaching the HLs and using them as the LOLT, in addition to facilitating the learning of additional languages such as English, the situation could change for the better.

Chapter 2 of the current study pointed out the continued emphasis placed by the Department of Basic Education on English FAL to the detriment of the status and development of the remaining school language subjects. This is short-sighted and unsound pedagogical reasoning. If all of the HLs are to represent the highest possible level of language use and ability, nine of these HLs cannot be relegated to a lower status and function than that accorded Afrikaans and English as languages of learning and teaching. The use of the HLs as media of instruction alongside other languages should receive priority as they fulfil an essential role in developing the desired level of additive bilingualism/multilingualism. Further to this, academic language proficiency is connected to the ability to think critically and engage in higher order reasoning (Mgqwashu 2011; Grosser & Nel 2013). If adequate knowledge of and ability to use the L1 or HL is essential for learning another language and developing critical thinking skills, much more needs to be done to develop and use all 11 HLs at the highest possible level. This too was recognised in the previous curriculum:

> If the teaching of literacy is firmly rooted in the world of the learners, they will make sense of their world and bring this knowledge to their reading of texts. Through reading, viewing and listening to texts, learners are able to reinterpret their world and use this knowledge to 'rewrite' their world in ways that contribute to social and environmental justice.

> For this to happen, learners must learn to listen, read and write in a language that is an integral part of their world. They will then be able to transfer the critical literacy they have acquired in their HL to their additional languages.
> (Department of Education 2008c: 13)

The shift in emphasis in CAPS away from indigenous languages that form an integral part of the world and life of students, and consequently away from using the HLs to scaffold and develop critical literacy, is lamentable. Although research studies on the possibility of employing multilingual pedagogies in which the HL is used in combination with other languages are still in an early stage, the preliminary indications are that in multilingual settings epistemological access can be facilitated through multilingual or bilingual pedagogies (Heugh 2014, 2015). In terms of this approach, more than one LOLT is being proposed. Nevertheless, irrespective of whether one or more LOLT is ultimately employed effectively in South African educational contexts, what is needed is a high level of ability in both the HL and the medium of instruction. In addition hereto, the intellectualisation argument has to go much further than using language for education. It must also extend to the mastery of different kinds of discourse: political, ethical, social, academic, occupational and aesthetic discourse, at least; hence the goals of the language learning curriculum in CAPS are laudable, if only they can be realised.

In order for South African students to have equal access to opportunities afforded by a democratic South Africa, the Departments of Basic and Higher Education need to devote far more attention to the findings of research on language learning and language policy in education, and integrate these in curriculum policy and practice. Unfortunately, however, scholars such as Balfour (2015) aver that in South Africa there is an insensitivity to research findings and a disregard of the need for students to develop sufficient language ability to operate in one language without the interference of another language, an aspect hailed by some educationists as the ultimate objective of language pedagogy (Balfour 2015: 189). It therefore remains uncertain as to when this situation is likely to change and the HLs will be prioritised as instrumental keys that unlock learning. Other ways may have to be devised to encourage the equitable development of the HLs.

7.4 The introduction of a common standardised examination component

In order to accommodate the sentiments expressed above, the format of the examination papers would need to be amended to make provision for a new examination component that would be comparable across all HLs. The viability hereof is discussed in the section that follows on the basis of insights gained from standardised tests used with success locally, as well as research findings on the development of parallel tests in more than one language.

Standardised tests that assess the readiness of students for higher education settings have been the focus of at least 50 years of research in the USA and UK and more than a decade in South Africa. Prominent tests such as the Test of English as a Foreign Language (TOEFL), developed by the Educational Testing Service (ETS) in the US, the Scholarly Aptitude Test (SAT), administered by the College Board in the US, and the Cambridge series of English proficiency tests that form part of the International English Language Testing System (IELTS), have specifically been designed to measure the language proficiency necessary for students to access higher education (Frantz, Bailey, Starr & Perea 2014). However, Frantz *et al.* point out that these tests have focused on the kind of language used in tertiary settings, and that suitable tests assessing language proficiency at school level are lacking. In South Africa too, the existing standardised tests such as the National Benchmark Tests (NBTs), Test of Academic Literacy Levels (TALL) and its Afrikaans derivative, Toets van Akademiese Geletterdheidsvlakke (TAG), are aimed at measuring the ability of prospective first-year university students to negotiate academic discourse at tertiary level (Cliff & Hanslo 2009; Van der Slik & Weideman 2010; ICELDA 2015). It would therefore make sense to design a specific test of language ability suitable for use as part of the Grade 12 school-leaving examination. One test that is undergoing development which could provide a useful blueprint for a common examination component is the Test of Advanced Language Ability (TALA), recently developed and piloted as part of a research initiative investigating the possibility of using item banking in the HL examination and developing parallel examination components (Steyn 2014).

7.4.1 The development of a Test of Advanced Language Ability (TALA)

TALA was inspired by the usefulness and high reliability of locally developed tests such as TALL and TAG.[41] These tests employ the original articulation of functional academic literacy deriving from the collaboration of academic literacy specialists involved in the Alternative Admissions Research Project (AARP) – the precursor of the National Benchmark Tests (NBTs) – and later refinements of the construct (Van Dyk & Weideman 2004a; 2004b). All versions of the tests have been validated and have produced combined *alpha* values well in excess of 0.9 across their administrations, testifying to their high reliability and usefulness as measurement artefacts that meet both the constitutive and regulative conditions for language tests expounded in Chapter 3 (cf. Weideman & Van der Slik 2008; Butler 2009; Van Dyk 2010). However, although some task types of TALL and TAG are suitable for inclusion in a Grade 12 common examination paper, a specific test of language ability would need to be designed for use at school level. Work has already commenced on the refinement of a Test of Advanced Language Ability (TALA) and pilot tests have been conducted in a selection of schools (Steyn 2014). Table 7.2 shows the envisaged blueprint with its five subtests.

Table 7.2: Outline of subtests and tasks in TALA (Weideman, Du Plessis & Steyn 2017)

Subtest and task type	Construct component(s) measured	Marks
Scrambled text	Cohesion and grammar; understanding relations between different parts of a text; sequence and order	5
Vocabulary knowledge	Vocabulary comprehension	10
Interpreting graphs and visual information	Understanding text type (genre); interpreting graphic and visual information; making distinctions; basic numerical computation	8
Text comprehension	Understanding metaphor and idiom; distinguishing between essential and non-essential information; classifying, categorising and handling data that make comparisons; extrapolating; synthesizing	25
Grammar and text relations	Vocabulary comprehension; textuality (cohesion and grammar); understanding text type (genre); communicative function	12

Each of the above subtests has been used reliably in standardised versions of TALL and TAG. For the purposes of piloting TALA, two parallel Afrikaans versions were developed. The first of these, the Toets van Gevorderde Taalvermoë (TOGTAV version 1), employed translation of TALA as test design technique, while the second version (TOGTAV 2), was developed

[41] Detailed information about TALL and TAG is available on the ICELDA website (ICELDA 2015).

independently on the basis of equivalent task types and construct similarity. All three versions have been piloted and found to be highly reliable, as indicated below.

Table 7.3: Reliability indices (Cronbach's alpha) of various TALA/TOGTAV pilots (Weideman, Du Plessis & Steyn 2017)

Version of test	Reliability (alpha)
TALA first pilot (187 items; n = 1244)	0.958
TOGTAV 1 first pilot (196 items; n = 368)	0.955
TOGTAV 2 first pilot (187 items; n = 357)	0.944
TALA (reduced 60-item version; n = 1244)	0.900
TOGTAV 2 (reduced 60-item version; n = 357)	0.831

One limitation of the pilot is that schools in the Bloemfontein area of central South Africa were used. Future pilots would need to include schools from different parts of the country in order to be able to generalise the results of the testing to a broader context. Nonetheless, the outcome of the initial piloting is encouraging. Even the shortened 60-item version of the tests has produced high reliability indices. The design of parallel versions of TALA in Sotho commenced recently and the team of researchers collaborating on the ICELDA project hope that ultimately tests will be developed and piloted in all the HLs. This will provide the best evidence for the usefulness of a common examination component.

In view of the objective to attain greater equivalence of standard across all HL papers, a common examination component such as TALA could provide a reliable benchmark and at the same time introduce a measure of fairness. In theory, it would also be possible to determine the concurrent validity of the remaining examination papers and subsections by using the TALA subtest scores.

South Africa could stand to benefit from a standardised test of advanced language ability at secondary school level in a number of ways. The design of the proposed examination component could form part of an initiative to strengthen the readiness of students for university study, as well as probe more adequately mastery of the higher order discourses, for example of political and ethical language as required by CAPS. The envisaged positive washback effect of the common paper could be supported by the development of materials that can be made available to students for independent study or that can be used by teachers as part of their language classroom activities. These materials could focus on inferential reading and advanced

language ability at FET level. This would require finances and political leverage and support, but if students can be afforded the opportunity to gain exposure to high-level language ability and analytical reasoning through the development of learning materials and standardised tests in the HLs, there could be considerable educational gains.

Since the existing examination papers are aimed at assessing differential language ability (as discussed in Chapter 4) through almost exclusively open-ended task types that provide little indication of scoring validity, it is proposed that the common standardised component should adopt the closed-ended multiple choice format used in TALA and that the emphasis be placed on generic language competence across various high-status discourse types.

The analysis of the selection of HL papers in Chapter 5 shows that only cursory attention has been given in past examination papers to the subabilities measured in TALA. One important aspect that needs assessing and that is missing from the current battery of examination papers is that of vocabulary. Knowledge of vocabulary and the polysemic nature of words is needed to access reading content (Van Wyk & Greyling 2008). The assessment of vocabulary should thus form part of the blueprint for the common paper. Although the current trend in language teaching and testing is to embed vocabulary in a full reading text so that the meaning of words is context dependent, there is still a need to include vocabulary tasks not related only to one particular text passage in order to assess a broader range of vocabulary. Provision can be made for testing vocabulary on the basis of lists of frequently occurring words (cf. Coxhead 2000; Flowerdew 2011) identified in texts. Here it is proposed that a corpus be compiled of vocabulary in the prescribed school language textbooks and a selection of vocabulary words common to all the HLs be made for the purposes of designing vocabulary items. The work already undertaken by the Centre for Text Technology (CText® 2015) can prove invaluable in this regard. CTexT® is located at the North-West University (NWU), Potchefstroom Campus, and runs a number of projects that facilitate translation and communication in the indigenous languages through advanced computer technology. Terminology databases in the official languages are also being developed. In addition, all the resources that have been developed are available free of charge, a further advantage of using the expertise that already exists.

Another aspect that has not been given prominence in the HL papers is the assessment of graphical and numerical literacy. Visual literacy tasks have been included in Paper 1, and visual texts also serve, somewhat unhappily, as ambiguous writing prompts in Paper 3, but the focus

has fallen on the aesthetic appeal of advertisements and photographs from the perspective of imagery, camera angles, fonts used, etc. Many of the visual literacy tasks analysed as part of the current study were found to be inappropriate indicators of language competence (see Chapter 5). The graphical and numerical literacy items contained in TALA, on the other hand, require the reading and interpretation of graphs to identify trends and distinguish between factual data expressed in numerical terms. These are abilities that indicate a high level of academic literacy and that would be appropriate in a test of advanced language ability.

Much mention is made in definitions of advanced language ability of the occurrence of complex grammatical constructions and the use of discourse markers (Frantz *et al.* 2014). Although Paper 1 includes a section that assesses grammatical aspects, more prominence should be given to the assessment of use of discourse markers and advanced features of language at HL level. It is proposed that the common paper also include tasks that assess syntactic connections, cohesion and coherence in an integrated manner such as in the case of TALA, specifically as these are assessed in its fifth and final subtest on grammar and text relations.

There are different ways to go about developing reliable and comparable tests in more than one language such as in the case of the piloted versions of TALA. These will be discussed next.

7.4.2 Translated parallel or independently developed construct-equivalent examination papers as design options

Standardised tests such as those designed by ICELDA, ETS and IELTS make use of parallel forms of tests developed through item banking of pre-piloted and refined test items. However, item banking may not be a viable option at school level in South Africa in view of security issues at HL examination centres, an aspect that could compromise the confidentiality of examination items. However, even if item banking is not used, the design of a common examination component each year by a team of examiners representative of the HL subjects and language testing professionals experienced in the objective assessment of language ability should be feasible.

There are at least three options that need exploring for the design of the common paper. Either one master copy can be generated by the team of examiners and language testing experts and

subsequently be translated into the HLs, or examination papers can be designed simultaneously per HL using the same construct and specifications, or thirdly, a combination can be used in which some parts are translated and others adapted. These options need to be piloted and the results analysed to see which would best meet the constitutive and regulative conditions discussed in Chapter 3. What is essential when employing translated or parallel test versions in cross-cultural assessment is that the test items must have the same meaning across all HL groups. The extent to which the different test versions can be considered equivalent involves determining whether the same construct is measured, and examining related parameters such as convergent validity and discriminant validity (for a full discussion see Van Dyk 2010, Chapter 7). This involves the production of correlation coefficients that examine patterns among different measures. Where the same construct is being measured, high correlations indicating a strong relationship should be produced.

Although a translated text cannot be considered to be identical to the original, translation has proven useful for teaching and testing purposes. For example, translation has played a key role in the intellectualisation of languages such as Japanese and Filipino (Alexander 2005), and should be explored to a greater extent for the development of the indigenous South African languages too. The translation of tests is also proving to be useful for the purposes of accountability in standards-based educational assessments and multilingual contexts (Sireci & Allalouf 2003; Stansfield 2003; Cronje 2009), but more research is needed to determine its viability in less commonly used languages such as Swati and Ndebele. A downside of translation is that it may not be suitable for all subject content and careful consideration would need to be given to the selection of texts. Nonetheless, translation is an archetypal form of inter-cultural communication by means of which peoples have been able to influence and develop one another for centuries. It can thus potentially be used as a mechanism for the intellectualisation of the indigenous languages (Alexander 2005: 6-7). In the opinion of Alexander (2005), translation is an effective cultural and social practice that makes possible "the invention of domestic literary discourses" that can increase the body of a language and its functional potential (Alexander 2005: 8). In any event, some texts used in HL examinations other than English or Afrikaans are already translated.

In Alexander's view if the indigenous languages are not developed and used as languages of learning and teaching, African countries will remain on the "periphery of the world economy" (Alexander 2005: 9). The promotion of indigenous languages in academia is considered by

those associated with the African Academy of Languages (ACALAN) – established in 2000 – as an important step to rectify ineffective educational systems on the continent and reduce intellectual and scientific dependence on Western nations. Very little investment has been made in developing the HLs in South Africa and the results thereof can be seen at universities where students struggle to complete their studies through the medium of English, drop-out rates continue to be high and through-put rates low (Du Plessis 2012a; Mhlongo 2014). It should be clear by now that the policy of using English as the only medium of instruction and assessment beyond the Intermediate Phase is not meeting the linguistic and cognitive needs of many students, and that we should be investigating bilingual or multilingual pedagogies. If the HLs are developed and intellectualised to a far greater extent at school level, this would enable a smoother transition to post-school education where more than one language could be used for teaching and testing purposes. Once the standard of the HLs is sufficiently high at school level and these languages are used to develop cognitive ability, we may start to see different results in higher education. More use should therefore be made of translation to generate academic materials in the HLs where these are lacking so as to facilitate their use in educational contexts.

Translation science is a discipline of its own and the translation of texts for the purposes of designing testing and teaching materials must involve suitably qualified persons (cf. Turkan & Oliveri 2014). The assistance of the South African Translators' Institute (SATI) and academics specialising in language testing would be essential to ensure that the designed examination artefacts comply with the principles of equivalence in both translation theory and test theory. Equivalence in translation theory does not signify sameness or identicality, but optimal similarity and refers to the relationship between source and target texts in respect of aspects such as content, style, text type, grammar and syntax, semantics and aesthetic features (Arffman 2010; Turkan & Oliveri 2014).

In language testing theory, equivalence is closely connected to reliability and validity and refers to the interchangeability of test versions that "measure the same construct in the same way" and are administered for the same purpose under the same conditions (Arffman 2010: 39). Of particular concern thus are measuring the same skills, covering the same content and having comparable levels of difficulty. Statistically speaking the test forms should be as close to identical as possible in respect of true score means, standard deviations and item difficulty (McQueen & Mendelovits 2003; Sireci & Allalouf 2003; Allalouf, Rapp & Stoller 2009;

Arffman 2010). Where test forms differ in error variance,[42] comparability is determined by means of test equation: this implies converting the scores from the respective test forms to a common scale and then correlating them with each other (Arffman 2010). Although complete equivalence is unattainable, it should still be possible to achieve a "relatively high level of equivalence (and validity)" (Arffman 2010: 37) when using translated texts and items. An important point here is that even if parallel or translated examination papers cannot be considered fully equivalent, for the cohorts of examinees writing the proposed new paper, it would be possible to compare and evaluate their performance more reliably and equitably per language group. This would still be preferable to the unacceptably high levels of subjectivity that characterise the scoring of the current papers, and the patently unfair outcomes of the HL final assessments.

In order to prevent construct irrelevant measurement errors in the planned examination component, it is proposed that factual texts on neutral topics be selected rather than texts rich in cultural idiom and metaphor that would pose translation complexities and potentially be biased against subgroups. Differential item functioning (DIF) or item bias is caused by differences in the item/construct relation among cultures and languages owing to translation errors or cultural/linguistic elements (Oliden & Lizaso 2013: 391). DIF is detected through methods such as IRT analyses and Mantel-Haenszel tests which identify test items that may be easier or more difficult for certain subgroups and that need to be investigated as possible sources of bias (Grisay & Monseur 2007; Van der Slik & Weideman 2010). Items that display undesirable forms of bias are usually replaced with other items or disregarded if not directly related to the construct under investigation. Turkan and Oliveri (2014) favour using both quantitative and qualitative methods of determining equivalence of items in multiple versions of a test. This view is similar to that of Van Dyk (2010: 204) who developed a matrix to evaluate construct representivity qualitatively in TAG and TALL items, in addition to using statistical indices. The concept scalar equivalence derives from psychological measurement where it is used to represent the highest possible form of equivalence between different test versions when the same construct is being measured and there is no item or measurement bias. Scalar equivalence enables valid inferences to be made when comparing averages across language groups and typically involves the conducting of t-tests and analyses of variance (Van de Vijver & Tanzer 2004; Ismail & Koch 2012).

[42] Error variance is a technical term used to indicate different types of measurement error (see Bachman 2004).

It is obvious from the above discussion that the development of multiple language versions of a common examination component would require time and resources. However, these costs would be offset by the savings offered by the electronic scoring of answer sheets and the usefulness of having a reliable benchmark for monitoring language learning at Grade 12 level. Not only would costs related to payment of markers be reduced considerably, but the results of the examination paper could be available within the span of a few days. Another advantage of this system is that it would be possible to carry out item analysis per HL paper after answer sheets have been processed electronically. As mentioned above and discussed in more detail in Section 3.4.2 of Chapter 3, item analysis is one way in which examination items can be checked for quality. At the moment the quality of the examination papers is determined through evaluations of a subjective nature – with reportedly inconclusive results – carried out by examiners and moderators. Item analysis provides an opportunity to provide an empirical and independent evaluation on the basis of factual data. For example, through statistical analysis of examination responses it will be possible to determine the reliability of each HL paper, compare item difficulty (cognitive challenge) and identify items that are biased towards a particular subgroup of examinees.

There are further ways of ensuring greater comparability. So as to attain the highest degree of equivalence possible, it is mandatory that the texts used in the examination papers, whether original documents or translated, should be comparable and the number of easy and difficult items should correspond. That would require more detailed item specifications. If after the analysis it should emerge that a particular paper had been far more challenging than the remaining papers were, the item statistics generated could provide a fair basis for statistically adjusting the marks of a cohort of examinees. A further benefit of objective measurement is that the results of the examination could serve as a barometer of language learning in schools across the country, similar to the way that current assessments such as the NBTs, TALL, TAG and TALPS provide a means of reflecting on the preparation of students for tertiary and postgraduate study (Du Plessis 2016).

Apart from meeting the constitutive requirements of validity and reliability, the proposed examination paper is also expected to comply with the juridical and ethical conditions that regulate language examinations. Fairness can be assured through the introduction of an examination component that provides similarity of task and cognitive challenge (relating to the

juridical mode). At the same time accountability can be reinforced through positive washback on the teaching side. The statistical analysis of results can also help to identify schools where additional resources may be required (with reference to the ethical dimension of testing). The latter forms an integral part of the notion of consequential validity alluded to in Chapter 3, and relates to the political and ethical modes of experience in relation to the technical instrument that is the assessment. After each administration of an examination the results should be interpreted to see whether the desired effect of the HL curriculum is being achieved and what further interventions may be necessary.

7.5 Accommodating the proposed common examination component

The aforegoing sections have demonstrated that there are sufficient reasons why the current examination papers need to be redesigned. Possible outlines for the revised papers are provided in the section that follows. Although the superordinate construct remains the same for the overall language examination, as delineated in Chapter 4, it should be kept in mind that the articulations of language ability will differ in accordance with the specific focus of each examination paper. Numerous design options are possible for the examination papers. However, if the primary objective is to attain greater equivalence of standard, only design options that contribute towards ensuring comparability of measurement across the HLs will be discussed.

7.5.1 Option 1: Adding an additional component to the existing examination papers

The existing format of the examination papers has remained the same for at least the past 10 years, which suggests that the examining authorities may show some initial reluctance towards introducing major changes in the way language ability is examined. The fact that Umalusi has commissioned a number of studies on the HL curriculum and examination is a positive step, but the decision to amend the papers would need to be taken at a higher level and would involve numerous stakeholders. If there is vehement opposition to changing the format of the papers, the addition of a fifth component, Paper 5 (TALA), could be negotiated. This would not interfere with the current system of examination and would at least enable the comparison to some extent of standards and performance across languages to some extent. However, this would not address any of the serious shortcomings identified in examination papers that were

discussed in the preceding chapters. Moreover, finding time to accommodate a fifth paper could prove to be difficult and would certainly increase the costs of the examination substantially. This option can therefore not be recommended.

7.5.2 Option 2: Replacing Paper 4 (Oral) with TALA

Another similar option, which would also be conservative but more cost and time effective, would be the retention of Papers 1-3, but the substitution of Paper 4 (Oral) with TALA.

Table 7.4: Removal of oral component

Formal assessment		
During the year	End-of-year examination	
25%	75%	
School based assessment (SBA)	End-of-year exam papers	
25% (100 marks)	62.5% (250 marks)	12.5% (50 marks)
• 1 test series in term 1 • 7 portfolio tasks • 2 examinations (mid-year exam/test & trial exam/test)	**Written examinations** Paper 1 (2 hours) – Language in context Paper 2 (2½ hours) – Literature Paper 3 (2½ hours) – Writing	**Paper 4: TALA** Advanced language ability in higher order discourses

Oral tasks could still be included in SBA, if deemed necessary. However, at HL level it can be argued that proficiency can be measured adequately up to Grade 10 level and that any further assessment of oral proficiency beyond that grade would be unnecessary. Apart from the fact that time could rather be made available for the development of other language abilities, such as advanced reading ability, it has already proven to be extremely difficult to moderate the assessment of oral ability (McCusker 2014). If the formal assessment of oral ability were to be removed, TALA could easily be reduced to 50 marks so as to retain the current combined examination mark of 300. This option would resolve the problem identified with double assessment of oral ability, both in SBA and as part of the external examination, and would enable comparability of standard in 16.6% of the examination and 12.5% of the combined SBA and external examination mark. Unfortunately, based on the discussions of the preceding chapters, this option would still be too limited and fall far short of addressing the main objective, which is to attain greater equivalence of standard and construct across HLs in Papers 1-3. Nonetheless, it could serve as a first step towards more comprehensive changes to the examination system later.

7.5.3 Option 3: Complete revision of all examination papers

Adopting as point of departure the fact that language ability relies on skill integration would enable a radical redesign of the papers to reflect an integrated communicative competence in the dominant material lingual spheres referred to in Chapter 4. At the same time the reliability of open-ended tasks could be increased by reducing construct-irrelevant items such as many of the essay writing items typically included in Paper 3. Shorter and more relevant writing tasks could be integrated with texts in the revised Paper 1 in order to increase task authenticity. This could enhance the reliability of scoring, as fewer choices would be provided to examinees and tasks would be more comparable across HLs in terms of complexity and cognitive demand.

Table 7.5: Complete revision of existing papers

Formal assessment	
School based assessment (SBA)	End-of-year exam examination
25% (100 marks)	75% (300 marks)
• 1 series of tests in term 1 • 7 portfolio tasks • 1 mid-year examination	**Written examinations** Paper 1 (2½ hours) – Integrated differential language ability (100 marks) Paper 2 (2 hours) – Literary and visual appreciation (100 marks) Paper 3 (1½ hours) – Advanced generic language ability (TALA) (100 marks)

On the SBA side, the number of formal assessments could be reduced to one test series in the first school term and one mid-year examination. The objective here would be to limit the amount of time devoted to "rehearsing" for the external examination and to provide more time for language learning. On the external examination side, the existing Papers 1 (Language in context) and 3 (Writing) could be combined into a single integrated paper that focuses on differential language ability and the processing of information. Paper 2 could be reduced to 2 hours and could include a task with a visual prompt. TALA could serve as Paper 3. This would substantially reduce the amount of examination time (6 hours compared to the current 7½ hours), and reduce the costs of marking the papers since Paper 3 would be marked electronically. Each of the three redesigned papers could contribute 100 marks. The common TALA type of examination paper would enable comparability of standard and construct in 33.3% of the examination and 25% of the combined SBA and examination mark. It would thus enable the reliable comparison of performance across languages to a far greater extent than any

of the previous options would. The proposed blueprints for the new examination papers are provided in the next section.

7.5.3.1 Redesigning Paper 1: Integrated differential language ability

A blueprint in the form of a table specifying the suggested number of tasks and items is provided below to show how the paper can be aligned better with the curriculum components and instruction time.

Table 7.6: Paper 1 – Integrated differential language ability

Section	Sub-abilities	Articulated sub-abilities[43]	Text specifications	Number of items and response formats	Mark allocation 100 marks	Time allowance 150 minutes
A	Processing Information	Careful reading of text Word recognition and interpretation Syntactic parsing Establishing propositional meaning at local and global level (i.e. within and across sentences) Inferencing Integrating information Communicating ideas Sharing views Constructing knowledge	1-2 reading texts with a combined word count of 700-800 (disjunctive orthography) or 500-560 (conjunctive orthography)[44]	15-20 items that require short responses and include a variety of item types (e.g. single word/phrase, full sentence, gap filling, multiple choice, etc.) 1 item that requires an extended response related to the reading text(s) Written response to be between 200-300 words (disjunctive orthography) or 150-200 words (conjunctive orthography)	50 marks in total Short response items (1-2 marks each) to contribute 30 marks Long response to contribute 20 marks	70 minutes
B	Differential ability to communicate at interpersonal and transactional levels	Communicative ability to produce a text (essay or letter) Text organisation Cohesion and coherence Register and style of communication Selection of appropriate language conventions	2 writing prompts in the form of short texts (choice of 2 essay topics or choice of 2 letter topics)	Examinees select 1 of 2 items provided and construct a written response of between 200-300 words (disjunctive orthography) or 150-200 words (conjunctive orthography)	20 marks	30 minutes
C	Knowledge of and ability to use appropriate language structures and conventions	Word classes, spelling, prefixes, suffixes, roots, figurative language, rhetorical devices, commonly confused words, collocations, linking words, abbreviations, acronyms, synonyms, antonyms, verb forms, auxiliaries, simple, compound, complex and compound complex sentences, active and passive voice, direct and indirect speech, concord, articles, infinitives, prepositions, punctuation.	1-2 texts	30 items that require short responses related to text(s) and that include different response techniques	30 marks (1 mark per item)	50 minutes

[43] These are based on the learning content expounded in CAPS (Department of Basic Education 2011a).

[44] For a discussion on disjunctive and conjunctive writing systems in the indigenous languages and how words are joined to other words grammatically in conjunctive orthographies see Taljard and Bosch 2006.

In order for the above format to be feasible, classroom instruction should follow the same integrated approach, as suggested earlier. Halliday (2007: 95), one of the scholars on whose work CAPS is based (as we have noted in Chapter 4), points out that "informal spoken language is every bit as systematic as formal written language", just organised differently. Learners learn through "many different registers, spoken and written, all at once". This supports an integrated approach to both classroom practice and test design. Meaning is made where discourse confronts experience, and it is discourse that "turns experience into knowledge" (Halliday 2007: 96). Social reality is thus constructed through discourse.

In the opinion of Halliday (2007), learning is predominantly a process that is driven lingually. Moreover, literacy applies specifically to writing and should be viewed as an activity rather than a form of knowledge (Halliday 2007: 99). Using written language should thus not be isolated from using language in another or in a more general sense. In other words, although separate concepts of literacy and oracy may exist, what is needed is a "unified notion of *articulacy*, as the making of meaning in language, in whatever medium". It is clear from what Halliday states that the principle of task authenticity in language teaching and testing requires that examinees should be able to engage with discourse before being asked to respond in writing. This also provides the theoretical justification for including integrated "reading" and "writing" tasks in Paper 1 as opposed to the current orthodoxy of a compartmentalised approach to developing reading and writing skills and assessing these separately.

For timed examination contexts, only certain genres are suitable for tasks that require writing lengthy responses. Agendas and recording of minutes should be covered as part of school-based assessment and based on authentic situations such as school assemblies, meetings of school committees, etc. In the same vein, speeches, dialogues and interviews cannot be executed purely as writing exercises and should also form part of continuous assessment at classroom level. Writing tasks such as articles that require extemporaneous knowledge and research should also not be included as examination items, unless an authentic format can be devised for their assessment. Amongst the genres that have been found to be appropriate for testing contexts, and that are suitable for the majority of examinees, are discursive essays, letters, proposals, reports and basic articles (Weir, Vidaković & Galaczi 2013: 239). The following examples from a selection of Cambridge English language examinations (Weir,

Vidaković & Galaczi 2013: 242-243) demonstrate the necessity of providing adequate writing prompts:

> **Example 1:**
> You have read the extract below as part of a newspaper article on the loss of national and cultural identity. Readers were asked to send in their opinions. You decide to write a **letter** responding to the points raised and expressing your own views. You **must** answer this question. Write your answer in **300-350** words in an appropriate style on the following pages.
>
> > 'We are losing our national and cultural identities. Because of recent advances in technology and the easy availability and speed of air travel, different countries are communicating more often and are therefore becoming more and more alike. The same shopping malls and fast food outlets can be found almost everywhere. So can the same types of office blocks, motorways, TV programmes and even lifestyles. How can we maintain the traditions that make each nation unique?'
>
> **Example 2:**
> You are employed as a researcher by your local tourist office. Your manager has asked you to write a proposal on how to attract more visitors, both from your own country and abroad, to your town or area. Within your proposal you should include ideas on how to improve the amenities in your town or area, and increase income from tourism. Write your **proposal**.
>
> **Example 3:**
> A monthly travel magazine has invited readers to contribute an article to a special edition entitled *The Best Way to Travel*. Write an article describing a memorable and enjoyable journey you have made and giving reasons for the means of transport used. Write your **article**.

In the above tasks, not only is the communicative event contextualised, but an indication is provided of possible power relations, and there is a clear sense of purpose and audience. It would therefore be possible to assess differential language ability across the material lingual spheres required in CAPS.

Although knowledge of grammar obviously plays a role in text comprehension (Section A of proposed new Paper 1) and the construction of texts (such as the example tasks provided above), there is still a need to include a separate section that focuses on grammatical aspects covered in the school curriculum. The HL examination serves as an assessment of the mastery of subject content, in addition to being an indicator of general language proficiency. Further reasons for assessing structural elements of language are the potential for high reliability, content validity and ease of scoring (Weir 2005: 172; Green 2014: 110). Nonetheless, the results obtained in Section C of the proposed redesigned Paper 1 should be considered representative of a particular component of the language curriculum, and not as an indication of overall proficiency.

7.5.3.2 Redesigning Paper 2: Literary and visual appreciation

The next section shows how some aspects of visual literacy can rather be incorporated in Paper 2 which covers the appreciation of a variety of literary texts. When the focus is on aesthetic aspects such as the effectiveness of camera angles, interpretation of images and appeal of fonts, etc., this represents a different kind of reading. The constructs being assessed differ vastly from those generally associated with language proficiency, which is why the visual literacy items included in Paper 1 were criticised in the content analysis carried out in Chapter 5. However, since students use multiple forms of communication and are regularly exposed to multi-media texts, they should have a basic knowledge of semiotics. Certain visual images can potentially also facilitate critical thinking skills (Piro 2002; Mostafa 2010; Moodley 2014). Because the reading or interpretation of aesthetic objects such as photographs and paintings serves as a "contextual intersection between literacy and culture" (Piro 2002: 128), it would make more sense if visual literacy is to remain part of the curriculum to include a visual literacy section in Paper 2 than in Paper 1. Careful consideration needs to be given to the selection of visual texts, though, as pointed out in Chapter 5. The selection of texts has proven to be highly problematic, because of the shortage of suitable materials in some of the HLs. The tendency to translate cartoons into the indigenous Bantu languages evident in the analysis of Sotho papers in the current study, is potentially problematic and should, if possible, rather be avoided. Just as jokes are virtually impossible to translate into another language without sacrificing the imaginative play on words and humour, so cartoons and other types of culturally-embedded texts are not ideal for translation. A further challenge is to find a reliable means of assessing visual literacy because the interpretation of images and text is multi-layered and of a highly subjective nature.

The proposed format for Paper 2 is provided on the next page.

Table 7.7: Paper 2 – Literary and visual appreciation

Section	Sub-abilities	Articulated sub-abilities[45]	Text specifications	Number of items and response formats	Mark allocation 100 marks	Time allowance 120 minutes
A	Visual appreciation	Understanding visual texts Discussing impact of layout, fonts, headings/captions, and camera angles Evaluating and responding to visual images Detecting bias	1-2 visual texts (photographs, advertisements, posters, illustrations)	5-10 items that require short constructed responses	10 marks in total (1-2 marks per item)	10 minutes
B	Poetry	Literal and figurative meaning Mood Theme and message Imagery Figures of speech, diction, tone, rhetorical devices, emotional responses, structure of poem (lines, stanzas, links, refrain), punctuation, repetition, sound devices (alliteration, consonance and assonance, rhyme, rhythm, onomatopoeia) and enjambment	4 seen poems 2 unseen poems	Examinees select 1 seen and 1 unseen poem Contextual questions (1-2 marks per item for a total of 10 marks per poem) should be provided	20 marks	30 minutes
Examinees to complete 1 essay question and 1 contextual question when responding to Sections C and D.						
C	Novel	Plot, sub-plot (exposition, rising action, conflict, climax, anti-climax, denouement/resolution, foreshadowing, flashback), setting, characterisation, role of narrator, messages, themes, ideologies, mood and tone, ironic twist/ending, timeline	1-2 text extracts per prescribed novel, not exceeding 300 words in total	2 tasks per prescribed novel Examinees select 1 essay task of 400-450 words[46] (40 marks) or answer 10-20 contextual questions (1-4 marks per item for a total of 40 marks).	40 marks	50 minutes
D	Drama	Plot, sub-plot (exposition, rising action, conflict, climax, anti-climax, denouement/resolution, foreshadowing, flashback), setting, characterisation, role of narrator, messages, themes, mood and tone, dramatic irony, ironic twist/ending, timeline, stage directions, link between dialogue/monologue/soliloquy and action	1-2 text extracts per prescribed drama, not exceeding 300 words in total	2 tasks per prescribed drama Examinees select 1 essay task of 250-300 words[47] (30 marks) or answer 10-20 contextual questions (1-3 marks per item for a total of 30 marks)	30 marks	30 minutes

[45] These are based on the learning content prescribed in CAPS (Department of Basic Education 2011a).
[46] Reduce for conjunctive orthography.
[47] Reduce for conjunctive orthography.

The respective sections of Paper 2 have been allocated according to the expected instruction time provided in CAPS. Most classroom time on the literature side is likely to be devoted to the novel, which is why Section C should contribute the most marks. The inclusion of a 10 mark visual literacy task is in the place of the third poetry task traditionally included.

7.5.3.3 Redesigning Paper 3: Advanced generic language ability

In the place of the problematic examination of writing ability in Paper 3, advanced generic language ability can be assessed reliably through a common HL examination component. The proposed blueprint for the common paper, which is based on the task and item types included in pilot administrations of TALA, follows on the next two pages.

Table 7.8: Paper 3 – Advanced generic language ability

Section	Sub-abilities	Articulated sub-abilities[48]	Text specifications	Number of items and response formats	Mark allocation 100 marks	Time allowance 90 minutes
A	Textuality	Knowledge of cohesion and coherence Understanding and responding to the communicative function of a text	1 text of about 100 words (disjunctive orthography) or 80 words (conjunctive orthography)	Scrambled text task 5 items (1 mark each) Multiple choice	5 marks	5 minutes
B	Interpreting graphs	Understanding texts in graphical format Numerical literacy Noticing trends Extrapolation and application of information Seeing sequence and order, and doing simple numerical estimations and computations that are relevant to academic information, that allow comparisons to be made, and can be applied for the purposes of an argument	1 graph	10 multiple choice items (1-2 marks each)	15 marks	15 minutes
C	Vocabulary	Advanced vocabulary knowledge Understanding and use of a broad range of vocabulary in context	20 sentences on different topics	20 multiple choice items (1 mark each)	20 marks	20 minutes
D	Understanding texts	Thinking critically, reasoning logically and systematically Distinguishing between essential and non-essential information, fact and opinion, propositions and arguments, cause and effect, and classifying, categorising and handling data that make comparisons Understanding the communicative function of various ways of expression in academic language (such as defining, providing examples, inferring, extrapolating, arguing)	1 text of about 600 words (disjunctive orthography) or 480 words (conjunctive orthography)	25 multiple choice items (1-3 marks each)	50 marks	40 minutes
E	Grammar and text relations	Understanding and responding to the communicative function of the text Knowledge of structural aspects of language	1 text of about 250 words (disjunctive orthography) or 200 words (conjunctive orthography	Modified cloze procedure with 10 multiple choice items (1 mark each)	10 marks	10 minutes

[48] These echo the stipulations for functional language use in CAPS, and are based on the task types used in the TALA pilot study (Steyn 2014), which in turn derive from those used successfully in tests such as TALL and TAG, and articulations of academic literacy provided by Patterson and Weideman (2013b).

The objective of the common paper would be to assess high-level language ability of a generic nature across all HLs in a manner that can be considered to have high construct validity and reliability and that will provide a fair basis for comparing performance. As reported in the analysis of examination papers in Chapter 5, there is little evidence in past examination papers of the assessment of advanced language ability. The proposed paper would allow 90 minutes for the reading of a selection of text types and the completion of 70 items. It would be challenging to complete within the required amount of time for students who are not proficient in the HL. The time constraint factor can be justified, however, on the basis that advanced language ability also implies an ability to read and process information fluently and hence in a short period of time (Bachman 1990: 123). This kind of ability would thus form an inherent part of the construct of advanced generic language ability.

Apart from enabling a distinction to be made between candidates of differing ability, the common paper could potentially also be used for placement purposes when admitting students to tertiary study. As pointed out in Section 7.3, students with a high level of ability in the HL are in a favourable position to transfer their language skills to the learning of other languages and subject matter in higher education contexts (Ellis 2008; Cummins 2011). They should be able to cope with the language demands of tertiary environments even if they proceed to study through the medium of a different language. This does not imply that the learning of the FAL is of less importance at school level. What is being advocated is the highest level of language ability attainable in both the HL and FAL, i.e. additive bilingualism/multilingualism at the highest possible level.

7.5.4 Option 4: Treating the language and literature components of the HL curriculum as separate school subjects

Apart from redesigning Paper 2 as proposed in the previous section, consideration should be given to separating the language and literature components of the HL curriculum. It may be argued that literary appreciation constitutes a different field of study to that of language learning and that the results of an examination such as Paper 2 should be reported separately. The present system of providing a unified score for all three HL papers is not useful to tertiary institutions who need to admit students to higher study, or in respect of organisations recruiting matriculants for the workforce. In both instances a

clear indication of language proficiency is required, albeit for different purposes. Unlike in the case of language assessments where results are reported on the basis of detailed descriptors, such as those used in the Common European Framework (CEFR 2013), the combined HL score per matriculant provides no indication of what kind of proficiency the graduating student has acquired. This is further exacerbated by the fact that the mark for Literature has been incorporated. Appendix E shows a few examples of CEFR descriptors used to indicate levels applicable to different language proficiency constructs. It is possible to report language performance in considerable detail per assessment opportunity. The results of the Grade 12 HL examination, on the other hand, are reported on the basis of seven perfunctory levels of achievement. The same levels are also used for reporting FAL results.

Table 7.9: Reporting of HL examination results (Department of Basic Education 2011a: 83)

Rating code	Description of competence	Percentage
7	Outstanding achievement	80-100
6	Meritorious achievement	70-79
5	Substantial achievement	60-69
4	Adequate achievement	50-59
3	Moderate achievement	40-49
2	Elementary achievement	30-39
1	Not achieved	0-29

As stated, the above cursory descriptors are used to admit students to higher learning. However, a distinction is not generally made by tertiary institutions between HL or FAL results. For example, in the Faculty of the Humanities at the University of the Free State a Grade 12 level 4 (50%) pass in the language of instruction is required for the general BA and BSocSc degrees, but no distinction is made between HL and FAL (University of the Free State 2016a: 86). In the case of the Faculty of Law (University of the Free State 2016b: 22), applicants for the mainstream LLB are required to have achieved a level 6 (70%) in Afrikaans or English, depending on which language will be the LOLT at university, but again no distinction is made between HL or FAL. In the Faculty of Natural and Agricultural Sciences, a generic level 4 (50%) pass in an official language is required (University of the Free State 2016c: 19). No cognisance is taken of the difference in curriculum content or level of ability between HL and FAL at school. The lack of

specificity here is already problematic but compounded by the fact that the mark obtained for Literature has been combined with that for language ability. There are allegations that the Literature component of the curriculum is highly repetitive and predictable, both in terms of prescribed works and examination questions. This is a further area that warrants investigation. Reporting the results separately would at least be helpful in terms of clarifying with greater certainty how proficient examinees are in a particular HL.

7.6 Conclusion

This chapter has discussed why the redesign of the current HL papers is necessary. Not only would this enable the inclusion of different task types with increased authenticity and reliability of measurement, but the incorporation of a common component that tests advanced language ability could have a positive washback effect and provide an impetus for the equitable development of all the HLs as academic languages. The potential benefits of being able to use the HL as the LOLT from Grade 1 to 12 are evident in the 2015 NSC examination results. The pass percentage of schools who offered Afrikaans at HL level and used it as the LOLT was 93.4% compared to the national pass percentage of only 70.7% in the instance of the remaining schools (SA Onderwysersunie 2016: 8). Considering the successes achieved in the NSC by students at Afrikaans-medium schools, it is worrying to see how – owing largely to political pressure – the number of schools using Afrikaans as the LOLT has decreased sharply within the span of only one year: from 267 in 2014 to 179 in 2015. The current arrangements, practices and influential political opinion are clearly having a dampening effect on school performance.

Table 7.10: Comparison of performance in Afrikaans-medium schools in 2014 and 2015 (SA Onderwysersunie 2016:9)

	2014				2015			
	TOTAL SCHOOLS	TOTAL WROTE	TOTAL ACHIEVED	%	TOTAL SCHOOLS	TOTAL WROTE	TOTAL ACHIEVED	%
GAUTENG	67	10 543	10271	97.4	43	8690	8583	98.8
KZN	5	320	316	98.8	6	339	326	96.2
LIMPOPO	1	258	256	99.2	1	222	222	100.0
MPUMALANGA	7	660	627	95.0	5	833	805	96.6
NORTHERN CAPE	26	1700	1395	82.1	17	1027	911	88.7
NORTH WEST	15	1833	1805	98.4	10	1185	1175	99.2
FREE STATE	30	2044	1984	97.1	17	416	395	95.0
EASTERN CAPE	27	2442	2073	84.9	21	2475	2009	81.2
WESTERN CAPE	89	9187	7782	84.7	59	6977	6316	90.5
TOTAL	267	28987	26509	91.5	179	22164	20742	93.4

The high pass percentage of students at Afrikaans-medium schools illustrates the possible gains to be made if other HLs are also developed as academic languages and used beyond the Foundation Phase as media of instruction. Obviously, English L1 students who study through the medium of English are in the same privileged position as those who are able to study through Afrikaans. The same opportunity needs to be created for the L1 speakers of the remaining official languages in order to argue the case for increased equivalence of standard in teaching and assessment.

In preparation for the proposed common examination paper, considerable resources would need to be developed in the previously disadvantaged Bantu languages to help students to expand their vocabulary and advance their language skills. The failure of so many undergraduate students to complete their degrees at universities in South Africa has been ascribed in part to “an articulation gap between school and higher education” (De Kadt 2015: 41), and the inadequate preparation of students (Lewin & Mawoyo 2014). Prioritising the HL component of the school curriculum and developing supportive literacy materials can play a pivotal role in equipping students to develop the language and cognitive abilities needed for immediate and eventual academic success.

In addition to the importance of developing the HLs to the same extent as Afrikaans and English, this chapter has shown how the redesign of the examination papers and

incorporation of a common examination component could ensure greater comparability of standard and increased fairness of measurement in at least one third of the HL examination. Further to this, it was illustrated how Papers 1-3 could potentially be redesigned and Paper 4 (Oral Assessment) be excluded. Increased authenticity of measurement would necessitate adopting an integrated approach to assessing reading and writing skills, and making provision for the assessment of both generic and differential language ability. As pointed out in Chapter 3, construct validity and high reliability indices provide the necessary justification for the interpretation and generalisation of test scores. Since a generic and differentiated language ability is the construct under consideration, an enriched, open view of language rather than the restrictive four skills-based approach should be pursued.

8 Findings and recommendations

8.1 Introduction

The research for this study was initiated in response to concerns raised by Umalusi that the HL examination papers are not of a comparable standard and that this disparity is reflected in the annual Grade 12 examination results. Extraordinarily high average marks are attained in some of the HLs and there are also alarming discrepancies per HL examination in some provinces. Further to this, when compared to non-language school subjects, the pass rates attained in the HL examinations are disproportionately high, which suggests that the examination papers are too easy. Although Umalusi has already commissioned several research projects to investigate these discrepancies and published the findings thereof, these studies can be described as being of a speculative nature and they do not offer solutions to address the problem. In fact, they merely serve to confirm perceptions that the papers are not of a similar standard and cognitive challenge.

By approaching the standard of the HL examination from a different angle to that adopted in the Umalusi studies referred to above, the researcher has attempted to identify dominant reasons for the discrepancies and developed a theoretical rationale that could enable the attainment of greater equivalence of standard across future HL examinations. In particular, three underlying causes for the disparities have been identified. The first of these pertains to the historical lack of parity of esteem of the HLs and their protracted inequitable treatment as academic languages. A second major reason for the varying standards of examination is the insufficient attention paid to the assessment of literacy and language ability in South Africa. This is evident in the lack of compliance with constitutive and regulative principles necessary for conducting language assessment in a fair and responsible manner and in the lack of conceptual clarity of constructs evidenced in the analysis of examination papers. The third contributing factor to the inequitable assessment relates to the format of the examination papers that does not support comparable assessment of language ability. On the basis hereof, the redesign of the entire HL examination is proposed.

In the section that ensues, the findings relevant to each of these three aspects will be provided and recommendations made on how greater equivalence of construct and standard can ultimately be attained in the HL examinations.

8.2 Lack of parity of esteem as academic languages

The indigenous Bantu languages are still marginalised in academe. This means that in order to create a comparable footing for the teaching and assessment of the HLs, they would need to be developed as academic languages. Chapter 2 discussed the historical context behind the teaching and assessment of the HLs and the advantageous position of English in particular. This section also covered the development of Afrikaans as an academic language in order to illustrate what can be achieved with the necessary political will and resources. Although the situation regarding the remaining official languages is far more complex and challenging than that of Afrikaans, it is sad that more than two decades into the new democracy hardly any progress has been made to develop the Bantu languages as academic languages. There is little sense in constantly revising curricula and standards without concerted practical measures to support the theoretical objective of attaining a high level of language ability in all HLs.

Recent attempts to encourage the use of languages other than English and Afrikaans (to at least some extent) at university level should be applauded and pursued. However, it defies logic to think that this is viable without first entrenching the academic status and use of these languages in primary and secondary schooling. Advocating an advanced register of language, as is evident from the official position as well as the policy documentation (CAPS) relating to this, presupposes the use of that language as a medium of learning and teaching. This is how a language becomes an academic one, develops that advanced register, and, once settled, maintains its position and status. The logical place to start developing an academic language is in the Foundation and Intermediate Phases of schooling from where the process can be expanded in secondary schooling and tertiary education.

The terminology used to designate the school language subjects is partly to blame for the diminished status. South Africans are predominantly bilingual and multilingual persons

who use different languages for different purposes, i.e. they are diglossic/multiglossic.[49] However, the use of the term HL suggests a low variety of language "acquired in the home and used in informal domains", even though CAPS explains that it is supposed to refer to the highest variety of language employed in the educational and other formal domains. It would thus be preferable to disregard the misleading wording that neither reflects the language situation in the home, nor captures the sense of a high-register language. This may also help to highlight the role of the languages in academic settings.

One possible way forward to break the impasse concerning the academic status of the HLs would be to move towards bilingual pedagogies at South African schools. To this end, far more attention would need to be devoted to the training and language skills of pre-service educators, one of the core issues raised at the first international Umalusi conference held in 2012. The researcher proposes that all aspirant educators should be required to continue studying HLs as part of their degree programmes. Education students should offer a minimum of two HLs and the language development modules should be compulsory for the entire duration of the undergraduate qualification. It takes time and effort to master a broad academic vocabulary and become an articulate and proficient user of a language. Some tertiary institutions require of Education students to complete only a basic course in one or more languages at first year level, as is the case, for example, at the University of the Free State (UFS). This is insufficient to address the poor language and literacy skills of many Education students, especially those who intend to teach in the critical Foundation and Intermediate Phases. Admittedly, students at the UFS who are planning to be language teachers *per se* in the Senior and FET phases are required to offer the chosen language subject up to second year level at least, but this may also be inadequate. The researcher is of the opinion that universities and other institutions of post-school learning are complicit in the neglect of the development of the HLs and that they should play a far greater role in both the increased use and standardisation of HLs for academic purposes.

The communicative competence of educators plays a key role in the transmission of knowledge and facilitation of learning, and the heuristic role of language at all levels of

[49] The term diglossia refers to the use of two varieties of a language for different purposes, or the use of separate languages for formal and informal purposes. See Ferguson, C. 1959. Diglossia. *Word*, 15: 325-40; Fishman, J. 1972. *The sociology of language.* Rowley, MA: Newbury House.

learning cannot be downplayed. If pre-service educators graduate with stronger language ability in more than one language and have been using at least two HLs as part of their training – and that over at least a three-year period – they should be better positioned to adopt bilingual pedagogies in school classrooms and facilitate the teaching and learning of subject content in more than one language.

To promote the sustained use and development of the HLs, it is further proposed that apart from using bilingual pedagogies consideration could be given to teaching some school subjects in one language and others in a second or additional language. This could promote a high level of functional bilingualism. Using languages other than Afrikaans and English as media of instruction would contribute towards parity of esteem and help ensure that the language diversity of learners is valued and harnessed as cultural and educational capital. At the same time, this would give effect to the rigorous implementation of the policy of additive bi-/multilingualism – which up to now has been misconstrued and applied erroneously. Rather than being strengthened, the HLs have all but been replaced by English. Unfortunately, the ideals of the 2008 curriculum to eradicate past inequalities on the language teaching side have not been realised, and what is more disturbing is that CAPS seems to disregard or at least to diminish the importance of learning through the medium of a first language or mother tongue. There is a definite policy shift towards a subtractive form of multilingualism which if allowed to continue can only impede the development of the indigenous Bantu HLs as academic languages.

Although English as the only LOLT has worked in some settings, predominantly where schools are located in well-resourced urban and middle class environments, this approach does not appear to be effective in schools that do not fit this description. On the whole the language ability and academic literacy levels of South African students entering tertiary training are lamentable, even though the orthodox approach up to now has been to employ English as the medium of instruction from the Intermediate Phase (or even earlier). It would appear that the intuitive assumption that the earlier English is introduced, the better it will be, is seriously flawed. In fact, it seems that the post-1994 Anglicisation of education has impeded the development of an academic meta-language in all HLs. Better results may have been achieved by encouraging learning and critical thinking in discourse conducted in languages other than English and Afrikaans from an early stage. Once again, the teachers play a central role in initiating classroom discussion and challenging their

students to delve deeper into the meaning of spoken and written texts. On this point, perhaps the time has come for introducing more stringent selection procedures for prospective teachers, in addition to increased investment in their language development and training. Dedicated and gifted educators are needed to teach with creativity and flexibility in multilingual and under-resourced schools in particular.

Chapter 2 also recognised the successes of the Department of Education in amalgamating 18 disparate education departments into a unified national department and introducing common curricula with centrally set examinations. Other achievements aimed at ensuring more equitable education include the increased amount of financing available for education and the funding of poorly resourced schools. The formation of quality control organisations such as the NQF and Umalusi were also commended as part of the quest of the new government to provide quality education for all students. In fact, the regular revision of curricula and standards bears testimony to the earnestness of the authorities to achieve this goal. Notwithstanding these accomplishments, however, the standard of the HL examination cannot be described as favourable and equitable, and it is clear that in terms of status English and Afrikaans are still in a stronger position than the remaining official languages. What seems to be missing is a realisation of the importance of the Bantu languages in supporting learning across all subject areas, in other words the potential educational gains when learners are given access to education in a language that is familiar and understandable to them.

The lack of benchmarking studies on the HLs on the part of Umalusi is regrettable. So far only English FAL has come under the spotlight. The failure to devote attention to the standard and use of the HLs as academic languages is one of the reasons why statistical adjustment of marks is needed to even out some of the disparities in examination results. In a brief comparison of language curricula elsewhere it is evident that international curricula such as those offered by Cambridge and the International Baccalaureate differentiate in terms of learning content and needs to a far greater extent than CAPS does when preparing learners for university study. This is an area that warrants further research, as the indications are that South African students at public schools are not prepared adequately for the tertiary environment. Further to this, the current level of English FAL also seems to be insufficient for university purposes (even though the existing Umalusi benchmarking studies indicate otherwise) and this is the level of English

provided to the majority of Grade 12 learners. Far more prominence needs to be given to the role of the school language subjects in preparing students for tertiary education if better results are to be achieved.

8.3 Lack of awareness of the importance of assessment principles

Language assessment has become increasingly sophisticated and specialised, yet the analysis of HL papers shows that this professionalism is not reflected in school language assessment in South Africa. There has been very little progress towards ensuring that the assessment of language ability as part of the NSC incorporates current views on the importance of valid and reliable forms of measurement. As a consequence, the Grade 12 HL examination varies in standard and cognitive complexity across HLs and the results can be considered neither fair nor credible. This runs counter to the objectives of eradicating all forms of inequality in education. The need for socially just and equitable forms of assessment is particularly pertinent in the light of the disparities emanating from previously unjust educational policies and programmes. Unfortunately, there does not seem to be the necessary awareness of the important role that constitutive elements such as validity and reliability fulfil in ensuring responsible and accountable assessment practices. The lack of cognisance of the social consequences of unfair assessment practices is also apparent.

8.3.1 The need to align language teaching and testing with current views in applied linguistics research

Chapter 3 provided a theoretical framework for the principled design of the HL papers to increase their credibility and usefulness as assessment artefacts. Different traditions of applied linguistics were referred to so as to elucidate how these continue to inform language teaching methodologies and testing practices. At the heart of applied linguistics research is the quest to devise responsible solutions to language-related problems. In the present study the issue to be addressed is how to attain greater equivalence of construct and standard in the HL examination, while at the same time being cognisant of current views and orthodoxies. The role of the NSC in managing linguistic diversity and supporting multilingualism in South Africa was foregrounded. Of relevance in this chapter is the continuity and reciprocity that was pointed out between successive

traditions of applied linguistics, and the principles contributed by each paradigm towards the responsible design of solutions relating to the teaching and testing of languages. Contemporary thinking in postmodern applied linguistics reflects consciousness of political power relations and highlights the need for political and social accountability in terms of the HL examination.

As a branch of applied linguistics, language testing has undergone its own paradigmatic shifts to reflect changing ideologies, oscillating between the pre-scientific/traditional approach and a modernist psychometric-structuralist design paradigm, from where the pendulum has swung to psycholinguistic-sociolinguistic, communicative and constructivist views on teaching and assessing languages. How greater attention can be given to the socially constructed nature of knowledge in formal testing contexts such as those applying to the HL examination is not clear. Postmodern language assessment leans towards an approach with focused eclecticism, and interrogates the use and consequences of testing. Of particular concern in the present study is the need for ecologically sensitive assessment that takes into consideration specific learning environments, and that does not rely on extemporaneous knowledge and privileged forms of literacy such as those identified in the analysis of examination tasks in Paper 3 (Writing).

The current orthodoxy on the language teaching and testing side remains the communicative approach with a high premium placed on authenticity of language tasks in real-world settings. However, we also see renewed interest in direct forms of assessment or what is referred to as the assessment of productive or expressive skills (tasks that involve mainly writing or speaking). These tasks that are reminiscent of the pre-scientific tradition of applied linguistics require subjective forms of assessment, but are complemented in highly acclaimed international large-scale language assessments with task combinations that can be subjected to psychometric analysis. In this way reliable measurement is ensured together with increased construct and content validity. This balance has not been achieved in the HL examination and there is a noticeable bias against objective forms of assessment. This bias is believed to derive partly from an inadequate awareness of the benefits of employing psychometric measurement and classical test or item response theory, and partly to insufficient training and qualifications in language assessment on the part of those tasked with the design and moderation of the HL papers.

At the same time, the offer of postgraduate qualifications in language testing at South African institutions needs to receive more prominence.

8.3.2 The need to apply principles that support the responsible design of the HL papers

The second part of Chapter 3 discussed the importance of essential design principles deriving mainly from the philosophical framework of Weideman (2017) while incorporating elements of the design approaches of Bachman and Palmer (1996), Weir (2005) and Green (2014). Together these provide a comprehensive basis for the responsible design of the HL examination papers. However, some of the design principles fell beyond the scope of the current study and require further research to do them justice. A decision was made to focus mainly on arguing the importance of applying the principles of validity, reliability and practicality to the HL examination within a postmodern paradigm that seeks to advance equitable and accountable assessment.

The Messick view of validity was contested in light of the fact that the process to validate an examination paper commences long before any scores are generated and the results interpreted. Obviously valid methods of scoring need to be employed and responsible inferences made on the ability shown by examinees, but the actual design of the examination paper needs to be adequate in the first instance. It was argued that if the construct for the HL paper is appropriate for the purposes of the examination and has been articulated and further specified adequately through suitably aligned examination tasks and items, only an illogical or unwarranted interpretation of scores would undermine the evidence of ability inferred from the examination results. By considering the social consequences of language assessment to form part of construct validity, those who ascribe to the Messick view exceed the theoretical definition of a construct. A decision was thus made to go beyond the orthodox notion of validity for the content analysis of examination papers and demonstrate how different kinds of validity all contribute useful information. As part hereof, validity evidence generated at different stages of the assessment system was discussed, i.e. *a priori* and *a posteriori* evidence. Of particular importance for the content analysis of examination papers was *a priori* theory-based and content/context forms of validity intrinsic to the condition of construct validity. The latter concerns the detailed articulation of the construct on the basis of accepted theories of language and

communicative competence and the language and cognitive processing involved in performing communicative tasks.

Although a third and middle phase of collecting validity evidence was included in the proposed theoretical framework, i.e. *per administratio* validity evidence pertaining to the actual examination event, this aspect was beyond the control of the study and was excluded. Nonetheless, it is a crucial part of ensuring equitable physical and environmental conditions for assessment. Only when all public schools have good and comparable infrastructure and physical resources can there be any mention of equitable examination settings that are conducive for assessment purposes.

The notion of reliability or consistency of measurement was referred to in detail to show its fundamental role in supporting the validity argument and enabling generalisation of inferences based on assessment results to non-testing domains, i.e. *a posteriori* evidence related conceptually to kinematic, numerical and physical analogies within the technically stamped domain of language assessment (Weideman 2009a, 2014). A distinction was made between statistically measurable properties of examination items (facility and discrimination values, distractor efficiency estimates and correlation coefficients) that contribute to consistency of measurement and scoring validity, and external factors that undermine reliability such as scoring systems that allow for considerable rater bias. Of the essential design principles, reliability is the most neglected quality of the HL examination. Even in a postmodern paradigm that values political and ethical considerations, there is a need to justify foundational aspects encapsulated in the validity of the construct and the reliability of the measurement.

The third important design principle discussed was that of practicality. This regulative idea contributes a technically disclosed form of validity and incorporates social and especially economic considerations that influence the design of the examination papers and may override, or suggest compromises in respect of other essential assessment qualities. If we consider how much teaching time is sacrificed for unreliable school-based and external assessment as well as the enormous financial investment that has to be made annually to cover the costs of administering the HL examination (with dubious results), it is obvious that time and cost efficient ways need to be found to conduct the assessment of the HLs.

The comprehensive framework proposed for the responsible and principled design of the HL papers also depicts the relation between the technical mode of experience and the normative conditions of validity and reliability at different stages of the assessment process. Only cursory reference was made in the study to the analogical reflections of the juridical, ethical and faith modes within the technical qualifying function of an assessment, all of which contribute other forms of evidence towards the usefulness of the HL examination. As pointed out, these are further areas of research that for the moment lie beyond the scope of the current analysis.

8.3.3 The need for conceptual clarity in respect of the construct used for measurement

The discussion of essential design principles in Chapter 3 laid the foundation for the definition and articulation of the HL examination construct in Chapter 4, an important part of ensuring theoretical defensibility on both the language teaching and testing side. The exploration of CAPS served to confirm adequate alignment of the curriculum with enriched sociolinguistic views on language and communicative competence. Three levels of mastery were identified: HL learners need to demonstrate proficiency at the social/basic, economic/professional and academic/educational level in order to meet the general aim of the curriculum to prepare learners for the workplace and further study. As far as the aims related specifically to the learning of the HLs are concerned, it is clear that advanced language skills are the distinguishing feature between HL and FAL level and that these far exceed what Cummins refers to as BICS. The construct of the HL examination would thus need to reflect a high level of ability in order to have validity.

Chapter 4 pointed out that the study of CAPS revealed a number of conceptual distinctions that should serve as the basis of an underlying construct for HL assessment. Language tasks needed to be defined functionally and should incorporate a variety of repertoires of language use. Learners should therefore be required to display both a generic ability to use language which demonstrates functional and formal aspects of language, as well as a highly differentiated capability peculiar to specific contexts and material lingual spheres. The HL examination tasks should be designed in such a way that examinees make language decisions from a repertoire of registers appropriate to a given

situation. In other words, they should have a chance to show knowledge and application of the norms and principles that typify language used in different contexts. The underlying construct for the HL examination papers was conceptualised as follows:

> The assessment of a differentiated language ability in a number of discourse types involving typically different texts, and a generic ability incorporating task-based functional and formal aspects of language. (Du Plessis, Steyn & Weideman 2013: 20)

Interestingly, the analysis of curriculum content showed that the dominant material lingual spheres covered in CAPS correspond closely with the approved non-language school subjects and this may provide an opportunity for making use to a greater extent of content based instruction (CBI) in the HL classroom. Further research is necessary to see whether HL classrooms are exposing learners to a range of discourse types and factual texts, or whether exposure is limited mainly to the study of the prescribed literature (aesthetic) texts.

All in all, the theoretical objectives and standards of CAPS appear to be in order for HL level and the curriculum seems to be comprehensive. One point of criticism is the separation of the traditional language skills in the curriculum document, even though their subsequent integration is encouraged. The researcher is of the opinion that CAPS can be redesigned to reflect different subsections that support an integrated, functional competence in more than one skill simultaneously (e.g. "Understanding Texts", rather than "Reading and Viewing"). Further to this, the fragmented and process approach to developing writing "skills" should be criticised as potentially counterproductive and impractical to execute in a pressurised school programme. It may help to limit the list of prescribed writing tasks to those that are indeed relevant to the world of the students and to remove those that require only basic communicative competence.

8.3.4 The need for item specification and construct representation

It is in Chapter 5 in particular that the issue of undesirable assessment practices comes to the fore. This chapter investigated the articulation of the underlying construct for HLs in a selection of English HL papers (2008-2012) by examining in detail the task specifications and marking memoranda. Each examination item of the two language papers, Paper 1 and Paper 3, was evaluated in terms of the *a priori* constitutive conditions

pertaining to the validity of construct, content and scoring. The analysis of Paper 2 (Literature) fell beyond the scope of the study and was excluded. A limitation of the study was the inability to obtain reliability coefficients from Umalusi since no statistical analysis of responses to examination items has been undertaken to date and the researcher was denied access to examples of scored examination papers. As a result hereof, only cursory attention could be given to *a posteriori* constitutive conditions for reliable measurement, and potentially subjective inferences had to be made.

Although the analysis of Paper 1 confirmed a measure of alignment with the teaching content of CAPS, too many factors undermined the validity of the examination. Content/context validity was compromised by the inclusion of unsuitable texts and a lack of item specifications. The reading comprehension tasks contained texts that were poorly written and covered topics that could potentially be biased in terms of culture and gender. Too many items were identified as problematic owing to poor formulation, repetition of parts of the item prompts in the memorandum answers, inappropriate answers in the memoranda and allowing copying from the reading passage where discussion was required. Far too many questions involved the expression of an opinion (just over half of the items analysed) resulting in an overrepresentation of this sub-ability, an aspect that weakened the construct validity of the paper. Item specification was also lacking in the summary writing task which could not be considered representative of the construct of *high* language ability. Theory-based validity was threatened by allowing summaries that did not reflect language and cognitive processing normally associated with summary writing. The fact that lifting phrases from the original text was allowed, examinees did not have to produce a coherent summary, and no penalties were applied for exceeding the prescribed length, are indicative of serious assessment shortcomings.

Another severe shortcoming in Paper 1 was the invalidity of scoring owing to a questionable system of mark allocation. Because of the lack of item specification, virtually no indication was given to examinees how marks would be earned and for the most part the memoranda allowed a blanket form of subjective scoring. There can be little mention of reliability of scoring where the acceptability of answers is left largely to the judgement of individual scorers of the examination scripts. This approach could in fact be interpreted by some critics as a deliberate strategy to ensure easy allocation of marks

and high pass rates. It constitutes poor assessment practice and is to be discouraged at all costs.

Section C, the last section of Paper 1, had a strong visual literacy rather than language focus. The inclusion of so many visuals in this part of the examination was criticised as being potentially unfair, since the construct of visual literacy requires cultural and extraneous knowledge. The content/context validity was also problematic owing to the fact that the analysis of visuals such as photographs and pictures does not replicate authentic language use in real social settings and forms part of the study of aesthetics rather than English. Across all years of analysis (2008-2012), items were identified that were problematic mainly owing to cultural or gender bias, poor formulation and deficiencies in the memoranda. On the whole, the individual items and suggested answers did not reflect higher order thinking, but lower order recall of information. Again, this weakens the construct validity of the paper as an assessment of high language ability.

Attention needs to be given, furthermore, to the structure and format of examination papers. Even if academic status and parity of esteem were to be achieved for all HLs, in addition to conceptual clarity on constructs and compliance with fundamental principles in language teaching and testing, some HL papers may still be more challenging than others. The analysis of a selection of Afrikaans, English and Sesotho papers in Chapter 6 revealed substantial variations in terms of the length of the papers and number of items to be completed, as well as a lack of measurement unit equivalence in respect of mark allocation. One relatively easy way to ensure greater technical equivalence would be to specify a minimum and maximum number of items per section and to apply measurement unit equivalence across all HL papers for greater consistency of measurement. Standardising the task specifications and providing fuller instructions to examinees would facilitate more equitable assessment. However, technical equivalence would mean very little without more radical steps to ensure greater comparability of constructs and standards across HLs.

In summary, the analysis of Paper 1 revealed a number of recurring features that undermined the validity and reliability of the examination across all five years of analysis. This points to a lack of cognisance of fundamental principles necessary for the responsible design of the papers, yet again highlighting the need for more attention to be paid to

competent and responsible assessment practices as part of the appointment and training of teams of examiners and moderators.

As far as the assessment of writing ability was concerned, the construct of Paper 3 exceeded that of HL ability by including imaginative ability and extemporaneous forms of knowledge, elitist or privileged forms of literacy that dominated colonial-era language examinations. Further to this, the creative writing tasks were not representative of texts learners would be expected to produce outside the school context and lacked situational authenticity. Another point of criticism was the way writing tasks were scored. The same rubric was used for different kinds of writing of varying levels of difficulty, rendering the system of scoring invalid. There was also overlap of content in the descriptors of the rubric and the distinctions between levels of writing were too fine. The complete absence of writing specifications made it impossible to assess differential writing ability and the application of normal conventions associated with discourse fields. In this way all forms of validity were compromised and a door was opened for test bias and measurement error.

Transactional or functional kinds of writing were assessed in Sections B and C of Paper 3. However, in the papers analysed, a lack of conceptual clarity on constructs was evident in the creative composition emphasis of many items in these sections. Some writing tasks were void of all authenticity and required the production of artificially contrived dialogues and interviews. In such cases there can be little evidence of theory-based or content/context validity. A further problem was that the writing tasks were of too basic and abridged a nature to provide any indication of ability at HL level. Even though Section C has since been combined with Section B, the kinds of tasks set remain of too elementary a level. On the scoring side, the same criticism of rubrics in Section A apply to Sections B and C in respect of overlap of content and inadequate distinction between levels of competence. One positive finding in the transactional writing section of Paper 3 was that fuller specifications were provided for some items.

In conclusion, it appears that some of the problems identified with Paper 3 are also attributable to assessment practices that are neither competent, nor responsible. The analysis of the three examination sections revealed a lack of conceptual clarity on the construct of writing, whereas the lack of clear item specifications obscured the important distinctions between different kinds of writing. Unfortunately, a brief comparison of

writing tasks and memoranda in a selection of Afrikaans and Sotho HL papers suggests that the problems identified in the English papers may occur in other HL papers as well. In the interests of equitable and comparable education, research projects on the remaining HLs should be prioritised. The findings of these studies could make a valuable contribution towards responsible assessment practices and enhancing the status and use of these languages, a prerequisite for equivalence on the teaching and assessment side.

8.4 Lack of systemic validity owing to the design of the HL examination

Alternative formats for the HL examination papers were proposed in Chapter 7 to increase the authenticity and reliability of assessment – essential constitutive conditions for foundational validity – and to promote the development of the HLs as academic languages – a necessary step for harmonising instruction and assessment and meeting regulative conditions (disclosed technical validity) that support the usefulness of the examination. There is a moral imperative to ensure that the HL examination has systemic validity, i.e. that the consequences of assessment are fair and beneficial to learners, and that the HL examination has a positive washback effect on teaching and learning. In their current format the HL papers fail to nurture the development of higher order thinking and differential language ability because they are designed to measure basic generic communicative ability (and that unreliably). Consequently, educators are likely to use past examination papers to coach their students for the school-leaving examination and neglect to devote attention to more advanced language skills. The result hereof is that even strong students may graduate with inadequate language proficiency and academic literacy to cope with the demands of tertiary education or the world of work. The current format of the HL examination needs revising if educational gains are to be made.

As a first step to alter the format of the examination, there should be closer alignment of the teaching and assessment programme in CAPS with the instructional time available and the objective of the curriculum to develop a high level of language ability. The analysis of syllabus weightings undertaken in Chapter 6 and the correlation of mark contributions of examination paper sections, showed that there is a mismatch. As much as 37.5% of the final mark in the HL examination derives from unreliable internal assessment at schools. The correlation of notional teaching and assessment hours showed

that proportionally too many marks are allocated in the external examination for the oral component of the curriculum, while too little instructional time is available to cover this skill. Chapter 6 also pointed out that proportionally speaking, too much separate instructional time is allocated for "Writing & presenting" as compared to the percentage weightings of these "skills" in terms of the external examination. This can be remedied by revising both the HL curriculum and examination to reflect an open and enriched view of language in which skills are naturally integrated. The overly ambitious separate oral and writing components of the curriculum can be incorporated with other learning content, instead of compartmentalising these skills. If speaking and writing are allowed to occur in conjunction with reading, viewing and thinking, the development of multiple "skills" can be facilitated simultaneously. This would be a more fruitful way to employ the limited teaching time available. At the same time, the format of HL Papers 1 and 3 should be amended to assess integrated ability and the mastery of higher order discourses: whereas Paper 1 could focus on differential language ability, Paper 3 could have as its focus generic language ability. By using open-ended and constructed-response task types in Paper 1 and closed-ended items in Paper 3, a balance could be achieved between subjective and objective marking – largely limiting the potential for undesirably high rater bias.

In a further attempt to increase the construct validity of the examination, brief suggestions were made for amending Paper 2 to incorporate visual literacy and semiotics with the assessment of literature. In this way, constructs with a cultural and aesthetic slant would be consolidated in a redesigned literary appreciation paper. It was pointed out that treating the language and literature components of the HL curriculum as separate school subjects and reporting the examination results of these individually could provide more useful information to potential employers of matriculants, as well as to institutions of further training tasked with selecting students for fields of study.[50] To date Paper 2 has not been validated in any research studies and there is no certainty as to the constructs being used or reliability of measurement, a further necessary area of study.

[50] It would be interesting to see how the results of the redesigned HL papers correlated with those of the annual benchmark tests (NBTs). Of course, the latter are only available in Afrikaans and English, an aspect that in itself may be considered unfair by some.

In order to introduce greater equivalence of construct and standard, it was further proposed that Paper 3 be introduced as a common examination paper using the validated construct of TALA as point of departure. In light of the fact that Umalusi reports have referred to difficulties experienced by teams of examiners to set higher order questions, a team of language testing experts should be tasked to design the papers annually in conjunction with HL subject specialists. In fact, this arrangement would be preferable for the design of Paper 1 as well if we consider the current lack of adherence to constitutive and regulative conditions for responsible assessment. As design options for a common paper, both translated parallel and independently developed construct-equivalent papers should be explored for this purpose. It was argued in Chapter 7 that such papers could support the development of the indigenous Bantu languages as academic languages since additional materials would need to be developed in the respective HLs to expose learners to vocabulary and language in a variety of discourse types and to introduce inferential types of questions that could stimulate cognitive thinking. Through the incorporation of statistical analyses it would be possible to monitor performance in Paper 3 reliably within and between HLs, and across years of assessment. This is essential as no measures are currently in place to support the reliability and generalisability of the results of the HL examination. A further possibility is that the results of the common paper could be used in a similar fashion to those of other standardised tests, for example those used by the International Association for the Evaluation of Educational Achievement (IEA). The Progress in International Reading Literacy Study (PIRLS) has been in use since 2001 to monitor reading achievement of Grade 4 students in at least 60 countries (IEA 2016). TALA could fulfil a similar purpose to PIRLS at Grade 12 level, just with a broader construct than reading achievement.

As part of the restructuring of the HL examination, it was suggested that the results for the Language and Literature components be reported separately. This would enable the use of bands of performance, which both quantify and qualify the abilities being measured. Certain bands may be associated with the ability to perform more challenging tasks (e.g. use language across the curriculum to analyse, apply and evaluate new information). Others may be associated with general communicative ability. At the moment the pass marks in HL subjects do not differentiate between different kinds of knowledge and mastery and there is too great an emphasis on basic communicative ability.

An essential part of the initiative to redesign the HL papers would obviously involve extensive consultation with the education authorities and teaching fraternity. A series of pilot tests and reception studies would be necessary to investigate the viability and acceptability of a common paper. As already mentioned, the further development and piloting of tests in the respective HLs would be essential to demonstrate how a common paper could supplement the existing papers to the benefit of all stakeholders.

Chapter 7 also alluded to the importance of prioritising the HLs from the early grades of learning as media of instruction in conjunction with languages such as Afrikaans and English in order to meet the objective of additive multilingualism and to facilitate the development of critical thinking skills. It is postulated that bilingual/multilingual pedagogies could foster increased epistemological access and are worth exploring. In this regard, endeavouring to ensure greater equivalence of standard in the HL examination by restructuring the format of the language papers represents only one rung of the ladder to educational and language justice and equality.

8.5 Conclusion

The focus of this study has fallen specifically on the way HL ability is assessed annually through the language examination. It is hoped that through the findings of the thesis, the important role played by the HL examination papers in reflecting standards and influencing learning will be recognised. In effect, the teaching and assessment of the HLs as school subjects serve as forms of situated language management. The proviso to use the HLs in higher register domains depends on their equitable development as academic and professional languages in tandem with sufficiently high standards of teaching and assessment. For real parity of esteem, both official status and comparability of use in a diversity of material lingual spheres is necessary. Equitability in assessment should thus be paralleled by equitability in learning opportunity. If the HL examination papers are amended so as to be more equivalent in all respects and future results continue to show disparities, this would signify unequal opportunities to learn deriving from other factors. Comparable school infrastructure, resources and teaching expertise are needed to complete the equation.

Comments related to competent and responsible assessment practices should not be interpreted as criticism of the subject knowledge of teams of examiners and moderators, many of whom are highly experienced and respected. However, by emphasising the need for the HL papers to comply with important principles such as validity and reliability, the researcher wishes to draw the attention of the examining authorities to the benefits to be gained from more sophisticated and responsible means of assessment. The latter would not only increase fairness and accountability of assessment, but would be fully in line with the intention of CAPS to achieve social transformation through redressing educational imbalances. The perception that the HL examination does not carry the same (lingual) currency as a few decades ago is attributable to the problematic design of the papers. Although the curriculum is clear enough, assessment practices warrant attention: not enough steps are in place to ensure the public accountability of the examination. It simply cannot be taken for granted that a formal qualification in language education and years of experience in teaching and setting examination papers necessarily guarantee assessment practices that are responsible, fair and accountable.

When viewed as a collective, the NSC results remain the best predictor of academic achievement at tertiary level. Although the HLs on their own may not have predictive validity, if we take into account their contribution towards the aggregate pass mark, it can be postulated that the more reliable the HL examination papers are, the more useful the overall NSC results are likely to be. Examination authorities need to show accountability to stakeholders by providing evidence that examination papers taken each year are close equivalents to those of previous years in terms of construct and cognitive challenge, and that across the different HL examination papers too there is a measure of equivalence.

In his seminal work on the history of language testing, Spolsky aptly sums up the efforts of applied linguists working in the field of modern language testing over many decades as follows:

> A central issue underlying this history has been the question of whether language proficiency can be measured on a definable dimension, like the time of a race or the distance of a jump or the number of goals, or whether it must be judged on a subjective set of criteria, like the performance of a diver, gymnast, or skater. But this dichotomy, like all other binary choices in two-valued logic, is almost certainly a false and unnecessary one. It is not too hard to realise that both sides of the argument are correct; that there are language abilities that are measurable, and that there are others that are only judgeable. (Spolsky 1995: 353)

The researcher expresses the hope that all stakeholders – and in particular education authorities, examiners and educators – would realise the near impossibility of measuring equitably what is only partially observable and immensely complex in a way that is acceptable and reasonable. By means of the proposed theoretical framework and principles for the responsible redesign of the HL examination, this analysis has made an earnest attempt to provide an impetus for the development of the HLs as academic and professional languages, and to ensure that in all of the HL examination papers the quantifiable is measured more objectively, and the immeasurable judged more responsibly and equitably.

Bibliography

About the International Baccalaureate®. [S.a.] [Online]. Available: http://www.ibo.org [2012, June 18].

Alderson, J.C. 2000. *Assessing reading.* Cambridge: Cambridge University Press.

Alderson, J.C., Clapham, C. & Wall, D. 1995. *Language test construction and evaluation.* Cambridge: Cambridge University Press.

Alderson, J.C. & Lukmani, Y. 1989. Cognition and reading: Cognitive levels as embodied in test questions. *Reading in a foreign language*, 5(2): 253-270.

Alexander, N. 2005. The potential role of translation as social practice for the intellectualisation of African languages. *Muratho* 5: 3-10. Keynote address delivered at the XVIIth World Congress of the International Federation of Translators held at Tampere, Finland, 4-7 August 2005.

Alexander, N. 2013a. *Thoughts on the new South Africa.* Auckland Park, South Africa: Jacana Media (Pty) Ltd.

Alexander, N. 2013b. *Language policy and national unity in South Africa/Azania.* E-book published by the Estate of Neville Edward Alexander. [Online]. Available: http://www.sahistory.org.za/archive/language-policy-and-national-unity-south-africa-neville-alexander [2014, April 7].

Allalouf, A., Rapp, J. & Stoller, R. 2009. Which item types are better suited to the linking of verbal adapted tests? *International journal of testing*, 9: 92–107. DOI: 10.1080/15305050902880686.

Anastasi, A. 1986. Evolving conceptions of test validation. *Annual review of psychology*, 3: 1-15.

Angoff, W. 1988. Validity: An evolving concept. In Wainer, H. & Braun, H. (Eds), *Test Validity.* Hillsdale, NJ: Erlbaum. Pp. 19-32.

Archer, A. 2008. Investigating the effect of Writing Centre interventions on student writing. *South African journal of higher education*, 22(2): 248-264.

Arffman, I. 2010. Equivalence of translations in international reading literacy studies. *Scandinavian journal of educational research*, 54(1): 37-59.

Association of Language Testing in Europe (ALTE). 2005. Materials for the guidance of test item writers. Language policy division, Council of Europe. [Online]. Available: http://www.alte.org/attachments/files/item_writer_guidelines.pdf [2013, August 14].

Bachman, L. & Palmer, A. 2010. *Language assessment in practice.* Oxford: Oxford University Press.

Bachman, L.F. 1990. *Fundamental considerations in language testing.* Oxford: Oxford University Press.

Bachman L.F. 2001. Designing and developing useful language tests. In Elder, C., Brown, A., Grove, E., Hill, K., Iwashita, N., Lumley, T., McNamara, T. & O'Loughlin, K. (Eds), *Experimenting with uncertainty: Essays in honour of Alan Davies*. Cambridge: Cambridge University Press. Pp. 109-116.

Bachman, L.F. 2004. *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.

Bachman, L.F. 2011. *How do different language frameworks impact language assessment practice*. Paper read at ALTE's Fourth International Conference, Jagiellonian University, KRAKÓW, 7-9 July.

Bachman L.F., Davidson, F., Ryan, K. & Choi, I.C. 1995. *An investigation into the comparability of two tests of English as a foreign language*. Cambridge: Cambridge University Press.

Bachman, L.F. & Palmer, A.S. 1996. *Language testing in practice*. Oxford: Oxford University Press.

Bachman, L.F. & Purpura, J.E. 2010. Language assessments: Gate-Keepers or door-Openers? In Spolsky, B. & Hult, F.M. (Eds), *The handbook of educational linguistics.* Oxford: Wiley-Blackwell. 456-468.

Bachman, L.F., Vanniarajan, A.K.S, & Lynch, B. 1988. Task and ability analysis as a basis for examining content and construct comparability in two EFL proficiency test batteries. *Language testing*, 5: 128-159.

Balfour, R.J. 2006. "A new journey with old symbols": University language policies, internationalism, multilingualism, and the politics of language development in South Africa and the United Kingdom. In Webb, V. & Du Plessis, T. (Eds), *The politics of*

language in South Africa. Studies in language policy in South Africa. Pretoria: Van Schaik. Pp. 1-14.

Balfour, R.J. 2015. *Education in a new South Africa: Crisis and change.* Cambridge: Cambridge University Press.

Berger, A.A. 1999. *Signs in contemporary culture: An introduction to semiotics.* Second edition. Salem, Wisconsin: Sheffield Publishing Company.

Bharuthram, S. & McKenna, S. 2012. Students' navigation of the uncharted territories of academic writing. *Africa education review*, 9(3): 581-594.

Bhorat, H. & Oosthuizen, M. 2008. Determinants of Grade 12 pass rates in the post-apartheid South African schooling system. *Journal of African economies*, 18(4): 634-666.

Blanton, L.L. 1994. Discourse, artefacts and the Ozarks: Understanding academic literacy. *Journal of second language writing*, 3(1): 1-16.

Bloor, T. & Bloor, M. 2013. *The functional analysis of English.* Third edition. New York: Routledge.

Blumfield, B.A. 2008. A historical view of the assessment of English HL at senior secondary school level in Kwazulu-Natal. Unpublished masters dissertation. Pretoria: University of South Africa.

Boughey, C. 2013. What are we thinking of? A critical overview of approaches to developing academic literacy in South African higher education. *Journal for language teaching*, 47(2): 25-41.

Brown, A. 2012a. Interlocutor and rater training. In Fulcher, G. & Davidson, F. (Eds), *The Routledge handbook of language testing.* New York: Routledge. Pp. 413-425.

Brown, H.D. 1994. *Principles of language learning and teaching.* Third edition. Englewood Cliffs: Prentice Hall.

Brown, J.D. 2012b. Classical test theory. In Fulcher, G. & Davidson, F. (Eds), *The Routledge handbook of language testing.* New York: Routledge. Pp. 323-335.

Brumfit, C. 1980. Being interdisciplinary – some problems facing applied linguistics. *Applied linguistics*, 1(2): 158-164.

BuaNews. 2012. South Africa spends more on education. 27 February. [Online]. Available: http://www.southafrica.info/about/education/budget2012-education.htm [2012, August 8].

Butler, H.G. 2007. A framework for course design in academic writing for tertiary education. Unpublished doctoral thesis. Pretoria: University of Pretoria.

Butler, H.G. 2009. The design of a postgraduate test of academic literacy: Accommodating student and supervisor perceptions. *Southern African linguistics and applied language studies*, 27(3): 291-300.

Cambridge International AS and A Level. [S.a.] [Online]. Available: http://www.cie.org.uk [2012, June 18].

Carstens, A. 2009. *The effectiveness of genre-based approaches in teaching academic writing: Subject-specific versus cross-disciplinary emphases.* Unpublished doctoral thesis. Pretoria: University of Pretoria.

Carstens, A. 2012. Using literacy narratives to scaffold academic literacy in the Bachelor of Education: A pedagogical framework. *Journal for language teaching*, 46(2): 9-25.

Casanave, C.P. 2012. Controversy and change in how we view L2 writing in international contexts. In Alsagoff, McKay, Hu & Renandya (Eds), *Principles and practices for teaching English as an international language*. New York: Routledge. Pp. 282-298.

Centre for Educational Testing for Access and Placement (CETAP). 2012. [Online]. Available: http://nbt.uct.ac.za/ [2012, 14 August].

Centre for Text Technology (CTexT®). 2015. [Online]. Available: http://www.nwu.ac.za/ content/ctextnews [2015, November 17].

Chaffee, J. 2014. *Critical thinking, thoughtful writing.* Sixth edition. Stamford, CT: Cengage Learning.

Chalhoub-Deville, M. & Deville, C. 2010. Nationally mandated testing for accountability: English language learners in the US. In Spolsky, B. & Hult, F.M (Eds), *The handbook of educational linguistics.* Chichester, West Sussex: Wiley-Blackwell. Pp. 510-522.

Chapelle, C.A. 2012. Conceptions of validity. In Fulcher, G. & Davidson, F. (Eds), *The Routledge handbook of language testing.* New York: Routledge. Pp. 21-33.

Chisholm, L. 2005. The state of South Africa's schools. In Daniel, J., Southall, R. & Lutchman, J. (Eds), *State of the nation: South Africa 2004-2005.* Cape Town: Human Sciences Research Council. Pp. 210-226.

CITO. 2005. *TiaPlus, classical test and item analysis.* Arnhem: Cito Measurement and Research Department.

Cliff, A. & Hanslo, M. 2009. The design and use of "alternate" assessments of academic literacy as selection mechanisms in higher education. *Southern African linguistics and applied language studies*, 27(3): 265-276.

Coe, R., Searle, J., Barmby, P., Jones, K. & Higgins, S. 2008. *Relative difficulty of examinations in different subjects: Summary report, July 2008*. CEM Centre, Durham University.

Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR). [Online]. Available: http://www.coe.int/t/dg4/linguistic/Cadre1_en.asp [2013, May 27].

Corder, S.P. 1973. *Introducing applied linguistics*. London: Penguin.

Coxhead, A. 2000. A new academic word list. *Tesol quarterly*, 34(2): 213-238.

Cronje, J.H. 2009. A systematic review of higher education admissions testing practices in Israel: Implications for South Africa. Unpublished masters dissertation. Port Elizabeth: Nelson Mandela Metropolitan University.

Crystal, D. 1981. *Directions in applied linguistics.* London: Academic Press.

Crystal, D. 1987. *The Cambridge encyclopedia of language.* New York: Cambridge University Press.

Culler, J. 1986. *Ferdinand de Saussure.* Ithaca, New York: Cornell University press.

Cumming, A. & Berwick, R. (Eds). 1996. *Validation in language testing.* Clevedon, UK: Multilingual Matters Ltd.

Cummins, J. 1979. Linguistic interdependence and the educational development of bilingual children. *Review of education research*, 49: 222-252.

Cummins, J. 2011. Bilingual and immersion programs. In Long, M.H. & Doughty, C.J. (Eds), *The handbook of language teaching.* Chichester, West Sussex: Wiley-Blackwell. Pp. 327-350.

Cummins, J. & Davison, C. (Eds). 2007. *International handbook of English language teaching Part 1.* New York: Springer Science+Business Media, LLC.

Cuvelier, P., Du Plessis, T. & Tech, L. (Eds). 2003. *Multilingualism, education and social integration. Studies in language policy in South Africa*. Pretoria: Van Schaik.

Davidson, F. & Lynch, B.K. 2002. *Testcraft: A teacher's guide to writing and using language test specifications.* New Haven: Yale University Press.

Davies, A. 2010. Accountability and standards.In Spolsky, B. & Hult, F.M. (Eds), *The handbook of educational linguistics*. Chichester, West Sussex: Wiley-Blackwell. Pp. 483-494.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. 1999. *Studies in language testing: Dictionary of language testing.* Cambridge: Cambridge University Press.

Davies, A. & Elder, C. 2005. Validity and validation in language testing. In Hinkel, E. (Ed.), *Handbook of research in second language teaching and learning*. Mahwah, New Jersey: Lawrence Erlbaum Associates. Pp 795-813.

De Kadt, E. 2015. The challenges of undergraduate education: Looking back and ahead. *The journal of the Helen Suzman Foundation*, issue 76, September 2015. [Online]. Available: http://hsf.org.za/resource-centre/focus/focus-76-the-idea-of-a-university/the-challenges-of-undergraduate-education-looking-back-and-ahead-elizabeth-de-kadt [2016, May 23].

De Kadt, J. 2006. Language development in South Africa – past and present. In Webb, V. & Du Plessis, T. (Eds), *The politics of language in South Africa. Studies in language policy in South Africa.* Pretoria: Van Schaik. Pp. 40-56.

De Saussure, F. 1974. *Course in general linguistics.* Glasgow: Fontana.

Department of Basic Education. 2010a. *The status of the language of learning and teaching (LOLT) in South African public schools: A quantitative overview*. Pretoria: Department of Basic Education.

Department of Basic Education. 2010b. *Report on the National Senior Certificate Examination Results*. Pretoria: Department of Basic Education.

Department of Basic Education. 2010c. *Education for all (EFA). 2009 Country report: South Africa*. Pretoria: Department of Basic Education.

Department of Basic Education. 2011a. *Curriculum and assessment policy statement: Grades 10-12 English HL*. Pretoria: Department of Basic Education.

Department of Basic Education. 2011b. *Curriculum and assessment policy statement: Foundation Phase Grades R-3*. Pretoria: Department of Basic Education.

Department of Basic Education. 2012a. Amendments to the examination guidelines for official HLs (HL); First Additional Languages (FAL) and History; National Senior Certificate (NSC). Circular E2 of 2012. [Online] Available: http://www.nwdesd.gov.za/images/principals/Circular%20E2%202012%20%282%29%20Assessment%20Framework.pdf [2015, October 19].

Department of Basic Education. 2012b. *Report on the National Senior Certificate Examination Results*. Pretoria: Department of Basic Education.

Department of Basic Education. 2012c. *National Senior Certificate Examination Schools Subject Report 2012*. Pretoria: Department of Basic Education. Pretoria: Department of Basic Education.

Department of Basic Education. 2012d. *National Senior Certificate Examination Technical Report 2012*. Pretoria: Department of Basic Education.

Department of Basic Education. 2012e. National protocol for assessment Grades 10-12. Government notice no. 1115 and no. 1116, Government Gazette no. 36042 of 28 December 2012. [Online]. Available: http://www.education.gov.za [2015, September 7].

Department of Basic Education. 2013. The incremental introduction of African languages in South African Schools: Draft policy. September 2013. [Online]. Available: http://www.education.gov.za/LinkClick.aspx?fileticket=DzmOFgWSE4E%3D&tabid=390&mid=1125 [2016, December 29].

Department of Basic Education. 2014. 39 day countdown to the start of the 2014 National Senior Certificate examination, 18 September 2014. Media release. [Online]. Available: http://www.education.gov.za/Newsroom/MediaReleases/tabid/347/ctl/Details/mid/1651/ItemID/3041/Default.aspx [2014, February 2].

Department of Basic Education. 2015a. Past examination papers. [Online]. Available: http://www.education.gov.za/Examinations/PastExamPapers/tabid/351/Default.aspx [2015, November 12].

Department of Basic Education. 2015b. *National policy pertaining to the programme and promotion requirements of the National Curriculum Statement Grades R-12*. [Online] Available: http://www.education.gov.za/LinkClick.aspx?fileticket=Rcf0UfEfk5s%3D&] [2015, February 2].

Department of Education. 1997. *Language-in-Education Policy 14 July 1997*. Published in terms of section 3(4)(m) of the National Education Policy Act, 1996 (Act 27 of 1996). *Government Gazette*, 18546: 19 December 1997.

Department of Education. 2001. *Language standardisation policy for official primary and official additional languages: Higher Grade/Standard Grade* (Government Gazette, 27 August 2001). Pretoria: Government Printer.

Department of Education. 2002. *Report on the Senior Certificate Examination*. Pretoria: Department of Education.

Department of Education. 2003. *Senior Certificate Examination report*. Pretoria: Department of Education.

Department of Education. 2004. *Senior Certificate Examination report*. Pretoria: Department of Education.

Department of Education. 2005. *The National Senior Certificate: A qualification at level 4 on the National Qualifications Framework (NQF).* Pretoria: Department of Education.

Department of Education. 2007. *National Curriculum Statement Grades 10-12 (General). Learning Programme Guidelines: Languages*. Pretoria: Department of Education.

Department of Education. 2008a. *National Curriculum Statement Grades 10-12 (General). Subject Assessment Guidelines. Languages: HL, First Additional Language, Second Additional Language*. Pretoria: Department of Education.

Department of Education. 2008b. *National Senior Certificate report.* Pretoria: Department of Education.

Department of Education. 2008c. *National Curriculum Statement Grades 10-12 (General) Learning Programme Guidelines Languages: HL, First Additional Language, Second Additional Language*. January 2008. Pretoria: Department of Education.

Desai, Z. 2003. Language-in-education policy in practice: A case study of a primary school in Khayelitsha, Cape Town. In Cuvelier, P., Du Plessis, T. & Tech, L. (Eds), *Multilingualism, education and social integration. Studies in language policy in South Africa*. Pretoria: Van Schaik. Pp. 141-162.

D'Oliveira, C. 2003. Moving towards multilingualism in South African schools. In Cuvelier, P., Du Plessis, T. & Tech, L. (Eds). 2003. *Multilingualism, education and social integration.* Studies in language policy in South Africa. Pretoria: Van Schaik. Pp. 131-140.

Dovey, T. 2010. Facilitating writing from sources: A focus on both process and product. *Journal of English for academic purposes*, 9: 45-60.

Du Plessis, C. 2012a. The design, refinement and reception of a test of academic literacy for postgraduate students. Unpublished masters dissertation. Bloemfontein: University of the Free State.

Du Plessis, C. 2012b. Facing up to literacy: Perceptions and performance in a test of academic literacy for postgraduate students. *Journal for language teaching*, 46(2): 123-139.

Du Plessis, C. 2014. Issues of validity and generalisability in the Grade 12 English HL examination. *Per Linguam*, 30(2): 1-19.

Du Plessis, C. 2016. Inferences from the Test of Academic Literacy for Postgraduate Students (TALPS). *Southern African linguistics and applied language studies*, 34(1): 1-16. DOI: 10.2989/16073614.2015.1108206.

Du Plessis, C. & Du Plessis, T. 2015. Dealing with disparities: The teaching and assessment of official languages at first language level in the grade 12 school-leaving phase in South Africa. *Language, culture and curriculum*, 28(3): 209-225. DOI:10.1080/07908318.2015.1083999.

Du Plessis, C., Steyn, S. & Weideman, A. 2013. Towards a construct for assessing high level language ability in Grade 12. Unpublished report submitted to Umalusi.

Du Plessis, C. & Weideman, A. 2014. Writing as construct in the Grade 12 HL curriculum and examination. *Journal for language teaching*, 48(2): 127-147.

Du Plessis, T. 2003. Multilingualism and language-in-education policy in South Africa – a historical overview. In Cuvelier, P., Du Plessis, T. & Tech, L. (Eds), *Multilingualism, education and social integration. Studies in language policy in South Africa*. Pretoria: Van Schaik. Pp. 99-119.

Du Plessis, T. 2006. Multilingualism at the South African Broadcasting Corporation (SABC) – perspectives on the SABC as co-orchestrator of language policy. In Webb, V. & Du Plessis, T. (Eds), *The politics of language in South Africa: Studies in language policy in South Africa series.* Pretoria: Van Schaik. Pp 82-99.

Educational Testing Services (ETS). 2012. The official guide to the TOEFL test. Fourth edition. New York: McGraw-Hill.

Eggins, S. 2005. *An introduction to systemic functional linguistics.* London: Continuum.

Elbow, P. 2000. *Everyone can write: Essays toward a hopeful theory of writing and teaching writing*. Oxford: Oxford University Press.

Ellis, R. 2008. *The study of second language acquisition*. Second edition. Oxford: Oxford University Press.

Fiske, E.B. & Ladd, H.F. 2004. *Elusive equity: Education reform in post-apartheid South Africa.* Washington DC: Brookings Institution Press.

Flowerdew, J. 2011. Corpora in language teaching. In Long, M.H. & Doughty, C.J. (Eds), *The handbook of language teaching.* Chichester, West Sussex: Wiley-Blackwell. Pp. 327-350.

Foley, L. 1990. Negotiating academic discourse. In Flower, L., Stein, V., Ackerman, J., Kantz, M., McCormick, K. & Peck, W.C. (Eds), *Reading-to-write: Exploring a cognitive and social process.* Oxford: Oxford University Press. Pp. 221-252.

Frantz, R.S., Bailey, A.L, Starr, L. & Perea, L. 2014. Measuring academic language proficiency in school-age English language proficiency assessments under new college and career readiness standards in the United States. *Language assessment quarterly*, 11(4): 432-457. DOI: 10.1080/15434303.2014.959123

Frederiksen, J.R. & Collins, A. 1989. A systems approach to educational testing. *Educational researcher*, 18(9): 27-32.

Fulcher, G. 2010. *Practical language testing.* London: Hodder Education.

Fulcher, G. 2015. *Re-examining language testing: A philosophical and social enquiry.* Abingdon, Oxon: Routledge.

Fulcher, G. & Davidson, F. 2007. *Language testing and assessment: An advanced resource book.* London: Routledge.

Fulcher, G. & Davidson, F. (Eds). 2012. *The Routledge handbook of language testing.* New York: Routledge.

García, O. 1997. Bilingual education. In Coulmas, F. (Ed.), *The handbook of sociolinguistics.* Oxford: Blackwell Publishers. Pp. 405-420.

Geldenhuys, J. 2007. Test efficiency and utility: Longer and shorter tests. *Ensovoort*, 11(2): 71-82.

Government Gazette No. 36797. 2013. *South African Qualifications Authority.* Notice Number 648, 30 August.

Grabe, W. & Kaplan, R. 1996. *Theory and practice of writing.* London: Longman.

Grabowski, J. 1996. Writing and speaking: Common grounds and differences toward a regulation theory of written language production. In Levy, C.M. & Ransdell, S. (Eds), *The science of writing.* New Jersey: Lawrence Erlbaum associates. Pp. 73-92.

Green, A. 2014. *Exploring language assessment and testing.* New York: Routledge.

Grisay, A. & Monseur, C. 2007. Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in educational evaluation*, 33: 69-86.

Grosser, M. & Nel, M. 2013. The relationship between the critical thinking skills and the academic language proficiency of prospective teachers. *South African journal of education*, 33(2): 1-17.

Guyer, R. & Thompson, N.A. 2011. *User's manual for Iteman 4.2*. St Paul Minnesota: Assessment Systems Corporation.

Habermas, J. 1970. Toward a theory of communicative competence. In Dreitzel, H.P. (Ed.), *Recent sociology 2.* London: Collier-Macmillan. Pp. 115-148.

Hall, C.J., Smith, P.H. & Wicaksono, R. 2011. *Mapping applied linguistics: A guide for students and practitioners*. New York: Routledge.

Halliday, M.A.K. 1978. *Language as social semiotic: The social interpretation of language and meaning.* London: Edward Arnold.

Halliday, M.A.K. 2007. *Language and education.* Volume 9 in the collected works of M.A.K. Halliday (Edited by Jonathan J. Webster). London: Continuum.

Hamp-Lyons, L. 1991. *Assessing second language writing in academic contexts.* Norwood, NJ: Ablex.

Hardcastle, P., Bolton, S., & Pelliccia, F. 2005. Test comparability and construct comparability across languages. In Taylor, L. & Weir, C.J. (Eds), *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity.* Proceedings of the ALTE Berlin Conference, May 2005. Cambridge: Cambridge University Press. Pp. 130-144.

Haupt, G. & Koch, E. 2012. The argument for evaluating monolingual language tests for equivalence across language groups. *Southern African linguistics and applied language studies*, 30(1): 65-76.

Hayes, J.R. 1996. A new framework for understanding cognition and affect in writing. In Levy, C.M. & Ransdell, S. (Eds), *The science of writing.* New Jersey: Lawrence Erlbaum Associates. Pp. 1-28.

Heugh, K. 2014. Turbulence and dilemma: Implications of diversity and multilingualism in Australian education. *International journal of multilingualism*, 11(3): 347-363. DOI: 10.1080/14790718.2014.921180.

Heugh, K. 2015. Epistemologies in multilingual education: Translanguaging and genre – companions in conversation with policy and practice. *Language and education*, 29(3): 280-285. DOI: 10.1080/09500782.2014.994529.

Hibbert, L. 2011. Language development in higher education: Suggested paradigms and their applications in South Africa. *Southern African linguistics and applied language studies*, 29(1): 31-42.

Hill, M. 1991. Writing summaries promotes thinking and learning across the curriculum: But why are they so difficult to write? *Journal of reading*, 34(7): 536-

539. Published by: International Reading Association. [Online]. Available: http://www.jstor.org/stable/40014578 [2013, May 28].

Hogan, T.P. 2007. *Psychological testing: A practical introduction.* Second edition. Hoboken, New Jersey: John Wiley & Sons.

Horn, G. 2006. Educational solutions to improve the employability of senior high school learners. *South African journal of education*, 26(1): 113-128.

Hughes, A. 2003. *Testing for language teachers*. Second edition. Cambridge: Cambridge University Press.

Hyland, K. 2002. *Teaching and researching writing.* London: Longman.

Hyland, K. 2003. Genre-based pedagogies: A social response to process. *Journal of second language writing*, 12(1): 17-29.

Hyland, K. 2009. *Teaching and researching writing.* Second edition. Applied Linguistics in Action Series (Edited by Candlin, C.H. and Hall, D.R.). New York: Taylor and Francis.

Hymes, D. 1972. On communicative competence. In Pride, J.B. & Holmes, J. (Eds), *Sociolinguistics: Selected readings.* Harmondsworth: Penguin. Pp. 269-293.

Inter-institutional Centre for Language Development and Assessment (ICELDA). 2015. Language test development and research. [Online]. Available: http://www.icelda.sun.ac.za/A [2015, November 2].

International Association for the Evaluation of Educational Achievement (IEA). 2016. TIMSS & PIRLS. [Online]. Available: http://timssandpirls.bc.edu/about.html [2016, March 6].

International English Language Testing System (IELTS). 2016. What is IELTS? [Online]. Available: https://www.ielts.org/ [2016, 6 March].

Ismail, G. & Koch, E. 2012. Investigating item and construct bias in an English verbal analogies scale. *Southern African linguistics and applied language studies*, 30(3): 325-338.

Jansen, J. & Taylor, N. 2003. *Educational change in South Africa 1994-2003: Case studies in large-scale education reform.* Country Studies Education Reform and Management Publication Series commissioned by the Education Reform and

Management thematic group of the World Bank, vol. II(1). [Online]. Available: http://www.jet.org.za/publications/research/Jansen%20and%20Taylor_World%20Bank%20report.pdf [2016, 2 February].

Jennings, M., Fox, J., Graves, B. & Shohamy, E. 1999. The test-takers' choice: An investigation of the effect of topic on language-test performance. *Language testing*, 16(4): 426-456.

Jetton, T.L. & Shanahan, C. (Eds). 2012. *Adolescent literacy in the academic disciplines: General principles and practical strategies.* New York: The Guilford Press.

John, V. 2012. Matric pass rate may be deceiving. *Mail & Guardian*. 6 January 2012. [Online]. Available: http://www.mg.co.za/article/2012-01-06-matric-pass-rate-may-be-deceiving [2012, 8 August].

Jones, J. 2012. Reliability and dependability. In Fulcher, G. & Davidson, F. (Eds), *The Routledge handbook of language testing.* New York: Routledge. Pp. 350-362.

Kamper, G. 2006. Implications of the position of indigenous languages in South Africa for the use of indigenous knowledge in community development. *Africanus*, 36(1): 75-87.

Kane, M.T. 1992. An argument-based approach to validity. *Psychological bulletin*, 112(3): 527-535.

Kane, M.T. 2004. Certification testing as an illustration of argument-based validation. *Measurement*, 2(3): 135-170.

Kaplan, R.B. (Ed.). 1980. *On the scope of applied linguistics.* Rowley, Massachusetts: Newbury House.

Kaschula, R. 2013. *Multilingual teaching and learning models at South African universities: Opportunities and challenges.* Seminar presentation, March 2013: NRF SARCHI Chair: Intellectualisation of African languages, multilingualism and Education, Rhodes University, Grahamstown.

Klos, M. 2011. The struggle for textual conventions in a language support programme. *Southern African linguistics and applied language studies*, 29(1): 55-65.

Koch, E. 2009. The case for bilingual language tests: A study of test adaptation and analysis. *Southern African linguistics and applied language studies*, 27(3): 301-317.

Krashen, S.D. & Terrell, T.D. 1983. *The natural approach: Language acquisition in the classroom.* Oxford: Pergamom.

Kumaravadivelu, B. 2006. *Understanding language teaching: From method to postmethod.* Mahwah, New Jersey: Lawrence Erlbaum Associates.

Kunnan, A.J. 2000. Fairness and justice for all. In Kunnan, A.J. (Ed.), *Fairness and validation in language assessment: Selected papers from the 19th Language Testing Research Colloquium, Orlando, Florida*. Cambridge: University of Cambridge Local Examinations Syndicate. Pp. 1-14.

Kunnan, A.J. 2010. Test fairness and Toulmin's argument structure. *Language testing*, 27(2): 183-189. DOI: 10.1177/0265532209349468.

Lado, R. 1961. *Language testing.* London: Longmans, Green and Co Ltd.

Lado, R. 1964. *Language teaching: A scientific approach*: London: McGraw-Hill.

Lafon, M. 2008. Asikhulume! African languages for all, a powerful strategy for spearheading transformation and improvement of the South African education system. In Lafon, M. & Webb, V. 2008. *The standardization of African languages: Language political realities*. Proceedings of a CENTREPOL workshop held at the University of Pretoria on March 29, 2007, supported by the French Institute of Southern Africa. Johannesburg: Institut Français d'Afrique du Sud. Pp. 34-59.

Larsen-Freeman, D. & Anderson, M. 2011. *Techniques and principles in language teaching.* Oxford: Oxford University Press.

Le, Phuong Loan. 2011. Assessing academic literacy of first year Vietnamese university students: How appropriate is the TALL? Unpublished masters dissertation. Groningen: University of Groningen.

Lewin, R. & Mawoyo, M. 2014. Student access and success: Issues and interventions in South African universities. Report published by Inyathelo: The South African Institute for Advancement, with the support of The Kresge Foundation. [Online]. Available: http://www.inyathelo.org.za/knowledge-services/inyathelo-publications/view-all-publications-for-download.html [2016, April 25].

Lightbrown, P.M. & Spada, N. 2006. *How languages are learned.* Third edition. Oxford: Oxford University Press.

Loggenberg, C. 2013. Motshekga grilled over matric exam markers. *Eyewitness news.* [Online]. Available: http://ewn.co.za/2014/02/18/Motshekga-grilled-over-matric-exam-markers [2015, October 14].

Louw, P.E. 2004. Anglicising postapartheid South Africa. *Journal of multilingual and multicultural development*, 25(4): 318-332.

Marrin, M. 2012. Forcing the great equality lie into school has squeezed learning out. *Sunday Times (UK)*, 8 April: 20-21.

McCusker, D. 2014. What's in the CAPS Package? A Comparative study of the National Curriculum Statement (NCS) and the Curriculum and Assessment Policy Statement (CAPS): FET Phase: English HL. Research cluster report presented to Umalusi on 11 June 2014.

McKay, S.L. & Rubdy, R. 2011. The social contexts of language teaching. In Long, M.H. & Doughty, C.J. (Eds), *The handbook of language teaching.* Chichester, West Sussex: Wiley-Blackwell. Pp. 9-25.

McNamara, T. 1996. *Measuring second language performance.* Harlow, Essex, UK: Addison Wesley Longman Ltd.

McNamara, T. 2000. *Language Testing.* Oxford: Oxford University Press.

McNamara, T. & Roever, C. 2006. *Language testing: The social dimension*. Oxford: Blackwell.

McQueen, J. & Mendelovits, J. 2003. PISA reading: Cultural equivalence in a cross-cultural study. *Language testing*, 20(2): 208-224.

Mda, T. 2000. Language in education. In Mda, T. & Mothata, S. (Eds), *Critical issues in South African education – after 1994.* Kenwyn, South Africa: Juta and Company Limited. Pp. 156-172.

Messick, S. 1980. Test validity and the ethics of assessment. *American psychologist*, 35: 1012-1027.

Messick, S. 1981. Evidence and ethics in the evaluation of tests. *Educational researcher* 10(9):9-20.

Messick S. 1988. The once and future issues of validity: Assessing the meaning and consequences of measurement. In Wainer, H. & Braun, I.H. (Eds), *Test validity.* Hillsdale, New Jersey: Lawrence Erlbaum Associates. Pp 33-45.

Messick, S. 1989. Validity. In Linn, R.L. (Ed.), *Educational measurement.* Third edition. New York: American Council on Education and Macmillan Publishing Company. Pp. 13-103.

Messick, S. 1995. Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American psychologist*, 50: 741-749.

Mesthrie, R. (Ed.). 2002. *Language in South Africa.* Cambridge: Cambridge University Press.

Mgqwashu, E.M. 2011. Academic literacy in the mother tongue: A pre-requisite for epistemological access. *Alternation*, 18(2): 159-178. ISSN 1023-1757.

Mhlongo, G.J. 2014. The impact of an academic literacy intervention on the academic literacy levels of first year students: The NWU (Vaal Triangle Campus) experience. Unpublished masters dissertation. Vanderbijlpark: North-West University.

Ministry of Education. 2003. The development of indigenous African languages as mediums of instruction in higher education. Report compiled by the Ministerial Committee appointed by the Ministry of Education, September 2003. [Online]. Available: http://www.education.gov.za/LinkClick.aspx?fileticket=VVy05Mi9bJY%3 D&tabid=452&mid=1036 [2014, 2 April].

Mncwago, J.B. 2015. An exploration of the discrepancy between classroom-based assessment and external summative assessment in English First Additional Language in Grade 12. Unpublished masters dissertation. Stellenbosch: Stellenbosch University.

Modisaotsile, B.M. 2012. *The failing standard of basic education in South Africa.* Africa Institute of South Africa: Policy Brief No. 72, March 2012.

Moodley, V. 2014. Quality and inequality in the assessment of visual literacy in Grade 12 examination papers across six South African languages. *Language matters: Studies in the languages of Africa*, 45(2): 204-223, DOI: 10.1080/10228195.2013.907331.

Mostafa, A.A. 2010. An exploration of teachers' integration of visual literacy in the Egyptian secondary language classrooms. *African journal of teacher education*, 1(1): 1-33.

Myburgh, J. 2015. The assessment of academic literacy at pre-university level: A comparison of the utility of academic literacy tests and Grade 10 Home Language results. Unpublished masters dissertation. Bloemfontein: University of the Free State.

Nel, C. & Kistner, L. 2009. The National Senior Certificate: Implications for access to higher education. *South African journal for higher education*, 23(5): 953–973.

Ockey, G.J. 2012. Item response theory. In Fulcher, G. & Davidson, F. (Eds), *The Routledge handbook of language testing.* New York: Routledge. Pp.336-349.

Oliden, P.E. & Lizaso, J.M. 2013. Invariance levels across language versions of the PISA 2009 reading comprehension tests in Spain. *Psicothema*, 25(3): 390-395.

Ouhalla, J. 1999. *Introducing transformational grammar: From principles and parameters to minimalism.* Second edition. New York: Oxford University Press.

Parker, F. 2012. Do the maths: Results not in line with SA's ambitions. *Mail & Guardian*. 10 January 2012. [Online]. Available: http://www.mg.co.za/article/2012-01-10-do-the-maths-matric-results-not-in-line-with-sas-ambitions/ [2012, 8 August].

Patterson, R. & Weideman, A. 2013a. The typicality of academic discourse and its relevance for constructs of academic literacy. *Journal for language teaching*, 47 (1): 107-123.

Patterson, R. & Weideman, A. 2013b. The refinement of a construct for tests of academic literacy. *Journal for language teaching*, 47(1): 125-151.

Pearson Test of English (PTE). 2016. [Online]. Available: http://pearsonpte.com/test-format/ [2016, 15 June].

Phillipson, R. & Skutnabb-Kangas, T. 2011. The politics of language teaching. In Long, M.H. & Doughty, C.J. (Eds), *The handbook of language teaching.* Chichester, West Sussex: Wiley-Blackwell. Pp. 26-41.

Piro, J. 2002. The picture of reading: Deriving meaning in literacy through image. *The reading teacher*, 56(2): 126-134.

Plakans, L. 2012. Writing integrated items. In Fulcher, G. & Davidson, F. (Eds), *The Routledge handbook of language testing.* New York: Routledge. Pp. 249-261.

Postma, M. & Postma, D. 2011. Who is laughing last in the South African classroom? A critical reflection on language in education. *Southern African linguistics and applied language studies*, 29(1): 43-54.

Pot, A. & Weideman, A. 2015. Diagnosing academic language ability: An analysis of the Test of Academic Literacy for Postgraduate Students. *Language matters: Studies in the languages of Africa*, 46(1): 22-43. DOI: 10.1080/10228195.2014.986665.

Pretorius, E.J. 2002. Reading ability and academic performance in South Africa: Are we fiddling while Rome is burning? *Language matters: Studies in the languages of Africa*, 33(1): 169-196. DOI: 10.1080/10228190208566183.

Prinsloo, J. 2002. Possibilities of critical literacy: An exploration of schooled literacies in the province of KwaZulu-Natal. Unpublished doctoral thesis. Johannesburg: University of the Witwatersrand.

Prinsloo, J. 2004. Examining the examination: The "worlding" of the matriculation language papers in KwaZulu-Natal. *Perspectives in education*, 22(1): 81-97.

Prinsloo, K. 2006. Towards a language strategy for Afrikaans: The process before, during and after the Afrikaans Language Indaba of 2004 in Stellenbosch. In Webb, V. & Du Plessis, T. (Eds), *The politics of language in South Africa: Studies in language policy in South Africa series.* Pretoria: Van Schaik. Pp 138-146.

Prinsloo, J. & Janks, H. 2002. Critical literacy in South Africa: Possibilities and constraints in 2002. *English teaching: Practice and critique*, 1(1): 20-38.

Rambiritch, A. 2012a. *Accessibility, transparency and accountability as regulative conditions for a post-graduate test of academic literacy*. Unpublished doctoral thesis. Bloemfontein: University of the Free State.

Rambiritch, A. 2012b. Challenging Messick: Proposing a theoretical framework for understanding fundamental concepts in language testing. *Journal for language teaching*, 46(2): 108-121.

Rambiritch, A. 2013. Validating the Test of Academic Literacy for Postgraduate Students (TALPS). *Journal for language teaching*, 47(1): 175-193.

Rambiritch, A. 2014. Towards transparency and accountability: The story of the Test of Academic Literacy for Postgraduate Students (TALPS). *Journal for language teaching*, 48(1): 71-87.

Read, J. 2010. Researching language testing and assessment. In Paltridge, B. & Phakiti, A. (Eds), *Continuum companion to research methods in applied linguistics.* London: Continuum International Publishing Group. Pp 286-300.

Reddy, V. (Ed.). 2006. *Marking matric*. Colloquium proceedings. Pretoria: Human Sciences Research Council Press. Pp xii-xix.

Republic of South Africa (RSA). 1995. *South African Qualifications Authority Act, 1995.* Act 58 of 1995. Pretoria: Government Printers.

Republic of South Africa (RSA). 1996a. *National Education Policy Act, 1996*. Act 27 of 1996. Pretoria: Government Printers.

Republic of South Africa (RSA). 1996b. *South African Schools Act, 1996.* Act 84 of 1996. Pretoria: Government Printers.

Republic of South Africa (RSA). 1996c. *Constitution of the Republic of South Africa, 1996*. Act 108 of 1996. Pretoria: Government Printers.

Republic of South Africa (RSA). 2008a. *General and Further Education and Training Quality Assurance, 2008.* Act No. 58. Pretoria: Government Printers.

Republic of South Africa (RSA). 2008b. *National Qualifications Framework Act, 2008. Act* No. 67 of 2008. Pretoria: Government Printers.

Riley, K.C. 2010. Language socialization. In Spolsky, B. & Hult, F.M. (Eds), *The handbook of educational linguistics*. Chichester, West Sussex: Wiley-Blackwell. Pp. 398-410.

Rosenberg, V.M. & Gabelnick, F.G. 1998. *Reading, writing, and thinking: Critical connections.* New York: Mcgraw-Hill.

SA Onderwysersunie/SA Teachers' Union (SAOU). 2016. Analysis of the National Senior Certificate results 2015. [Online]. Available: http://www.saou.co.za/wp-content/uploads/2016/01/ANALYSIS-OF-THE-NSC-RESULTS-2015.pdf [2016, 16 March].

Saville, N. 2012. Quality management in test production and administration. In Fulcher, G. & Davidson, F. (Eds), *The Routledge handbook of language testing.* New York: Routledge. Pp. 395-412.

Sealey, A. & Carter, B. 2004. *Applied linguistics as social science.* New York: Continuum.

Searle, J.R. 1969. *Speech acts: An essay in the philosophy of language.* London: Cambridge University Press.

Shin, D. 2012. Item writing and writers. In Fulcher, G. & Davidson, F. (Eds), *The Routledge handbook of language testing.* New York: Routledge. Pp. 237-248.

Shohamy, E. 2001a. *The power of tests: A critical perspective on the uses of language tests.* Harlow, UK: Longman.

Shohamy, E. 2001b. Democratic assessment as an alternative. *Language testing*, 18(4): 373–391.

Shohamy, E. 2006. *Language policy: Hidden agendas and new approaches*. New York: Routledge.

Sireci, S.G. 1997. Problems and issues in linking assessments across languages. *Educational measurement: Issues and practices*, 16(1): 12-19.

Sireci, S.G. & Allalouf, A. 2003. Appraising item equivalence across multiple languages and cultures. *Language testing*, 20(2): 148-166.

Slomp, D.H. 2005. Teaching and assessing language skills: Defining the knowledge that matters. *English teaching: Practice and critique*, 4(3): 141-155.

Snow, C.E. & Uccelli, P. 2009. The challenge of academic language. In Olson, D.R. & Torrance, N. (Eds), *The Cambridge handbook of literacy.* Cambridge: Cambridge University Press. Pp. 112-133.

Solidarity Research Institute. 2012. The South African labour market and matriculants' prospects. [Online]. Available: http://www.moneyweb.co.za/mw/action/media/downloadFile?media_fileid=17832 [2014, June 19].

Solidarity Research Institute. 2015. Matric report: The South African labour market and the prospects for the matriculants of 2015. [Online]. Available: http://www.solidariteit.co.za/wp-content/uploads/2016/01/Matric_Report_2015.pdf [2017, January 18].

Spaull, N. 2012. Education in SA: A tale of two systems. [Online]. Available: http://www.politicsweb.co.za/politicsweb/view/politicsweb/en/page71619?oid=323272&sn=Detail [2014, February 19].

Sperling, M. 1996. Revisiting the writing-speaking connection: Challenges for research on writing and writing instruction. *Review of educational research*, 66(1): 53-86.

Spolsky, B. 1995. *Measured words*. Oxford: Oxford University Press.

Spolsky, B. 2012. Language testing and language management. In Fulcher, G. & Davidson, F. (Eds), *The Routledge handbook of language testing.* New York: Routledge. Pp. 495-505.

Spolsky, B. & Hult, F.M. (Eds). 2010. *The handbook of educational linguistics*. Chichester, West Sussex: Wiley-Blackwell.

Stansfield, C.W. 2003. Test translation and adaptation in public education in the USA. *Language testing*, 29(2): 189-207.

Statistics South Africa. 2012. *Census 2011 Census in brief*. Pretoria: Statistics South Africa.

Steyn, S. 2014. *Towards the development of equal and fair HL assessments: Outline of a pilot study*. Interim report submitted to Umalusi.

Stockmeyer, N.O. 2009. Using Microsoft Word's readability program. *Michigan bar journal*, 88: 46-47. [Online]. Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1210962 [2016, June 1].

Stoynoff, S. 2009. Recent developments in language assessment and the case of four large-scale tests of ESOL ability. *Language teaching: Surveys and Studies*, 42(1): 1-40.

Suomela-Salmi, E. & Dervin, F. (Eds). 2009. *Cross-linguistic and cross-cultural perspectives on academic discourse.* Amsterdam: John Benjamins Publishing Company.

Taljard, E. & Bosch, S.E. 2006. A comparison of approaches to word class tagging: Disjunctively vs. conjunctively written Bantu languages. *Nordic journal of African studies*, 15(4): 428-442.

Taylor, L. & Weir, C.J. 2008. *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity*. Proceedings of the ALTE Berlin Conference, May 2005. Cambridge: Cambridge University Press.

The Economist. 2012. Education in South Africa still dysfunctional: Standards still leave a lot to be desired. Author unknown. 21 January edition. [Online]. Available: http://www.economist.com/node/21543214 [2014, February 5].

The Times. 2013. Wise to be cautious in introducing third language teaching. Editorial. 7 August 2013. [Online]. Available: http://www.timeslive.co.za/opinion/editorials/2013/08/07/wise-to-be-cautious-in-introducing-third-language-teaching [2014, April 7].

Truscott, J. 1996. Review article: The case against grammar correction in L2 writing classes. *Language learning*, 46(2): 327-369.

Tshotsho, B.P. 2013 Mother tongue debate and language policy in South Africa. *International journal of humanities and social science*, 3(13): 39-44.

Turkan, S. & Oliveri, M.E. 2014. Considerations for providing test translation accommodations to English language learners on common core standards-based assessments. Educational Testing Systems (ETS) research report ETS RR-14-05. [Online]. Available: http://wileyonlinelibrary.com [2015, November 27].

Uccelli, P. & Snow, C. 2010. A research agenda for educational linguistics. In Spolsky, B. & Hult, F.M. (Eds), *The handbook of educational linguistics*. Chichester, West Sussex: Wiley-Blackwell. Pp. 626-642.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2004a. *Approaches to quality assurance in the GET and FET bands.* Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2004b. *Investigation into the standard of the Senior Certificate examination.* Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2006. *Quality Assurance of assessment: policies, directives, guidelines and requirements.* Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2007. *Making educational judgements: Reflections on judging standards of intended and examined curricula.* Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2008a. *The role of IRT in selected examination systems.* Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2008b. *Learning from Africa.* Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2008c. *Signalling performance: An analysis of continuous assessment and matriculation examination marks in South African schools.* Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2009a. *2008 Maintaining standards report part 1: Overview*. Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2009b. *2008 Maintaining standards report part 2: Curriculum evaluation.* Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2009c. *2008 Maintaining standards report part 3: Exam paper analysis.* Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2010. *Evaluating the South African National Senior Certificate in relation to selected international qualifications: A self-referencing exercise to determine the standing of the NSC.* Joint research project undertaken by Umalusi and Higher Education South Africa (HESA). Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2011a. *Understanding statistical moderation.* Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2011b. *All the cattle in the kraal: An overview of Umalusi's research 2003-2011*. Research report. Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2011c. *Comparative Analysis of the National Senior Certificate HL, 2008-2010: Afrikaans; English; isiNdebele; isiXhosa; isiZulu; and SiSwati*. Unpublished research report. Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2012a. *The standards of the National Senior Certificate HL examinations: A comparison of South African official languages.* Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2012b. *Developing a Framework for assessing and comparing the cognitive challenge of HL examinations*. Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2012c. [Online]. Available: http://umalusi.org.za.html [2012, June 6].

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2012d. *Technical report on the quality assurance of the examinations and assessment of the National Senior Certificate (NSC) 2012*. Pretoria: Umalusi.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2012e. *Standards in Education and Training: The challenge*. Umalusi 2012 conference proceedings.

Umalusi (Council for Quality Assurance in General and Further Education and Training). 2013. *Directives for certification National Senior Certificate (schools)*. Pretoria: Umalusi.

University of the Free State. 2016a. Rule book 2016 Faculty of the humanities: Undergraduate programmes. [Online]. Available: http://apps.ufs.ac.za/dl/yearbooks/259_yearbook_eng.pdf [2016, May 11].

University of the Free State. 2016b. Undergraduate faculty rules: Law. [Online]. Available: http://apps.ufs.ac.za/dl/yearbooks/266_yearbook_eng.pdf [2016, May 11).

University of the Free State. 2016c. Faculty of natural and agricultural sciences rule book 2016. [Online]. Available: http://apps.ufs.ac.za/dl/yearbooks/262_yearbook_eng.pdf [2016, May 11].

University of the Free State. 2016d. Faculty of health sciences. [Online]. Available: http://www.ufs.ac.za/docs/librariesprovider25/all-documents/faculty-booklet-1007-eng. pdf?sfvrsn=0 [2016, October 24].

Vähäpässi, A. 1982. On the specification of the domain of school writing. In Purves, A.C. & Takala, S. (Eds), *An international perspective on the evaluation of written composition.* Oxford: Pergamon. Pp. 265-289.

Van de Vijver, F. & Tanzer, N.K. 2004. Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de psychologie appliquée*, 54: 119–135. DOI:10.1016/j.erap.2003.12.004.

Van de Vijver, F.R.J. & Leung, K. 1997. *Methods and data analysis for cross-cultural research*. California: Sage Publishers.

Van der Slik, F. & Weideman, A. 2005. The refinement of a test of academic literacy. *Per linguam*, 21(1): 23-35.

Van der Slik, F. & Weideman, A. 2007. Testing academic literacy over time: Is the academic literacy of first year students deteriorating? *Ensovoort*, 11(2): 126-137.

Van der Slik, F. & Weideman, A. 2008. Measures of improvement in academic literacy. *Southern African linguistics and applied language studies*, 26(3): 363-378.

Van der Slik, F. & Weideman, A. 2009. Revisiting test stability: Further evidence relating to the measurement of difference in performance on a test of academic literacy. *Southern African linguistics and applied language studies*, 27(3): 253-263.

Van der Slik, F. & Weideman, A.J. 2010. Examining bias in a test of academic literacy: Does the Test of Academic Literacy Levels (TALL) treat students from English and African language backgrounds differently? *Journal for language teaching*, 44(2): 106-118.

Van der Walt, C., Evans, R. & Kilfoil, W.R. 2009. *Learn 2 teach. English language teaching in a multilingual context.* Fourth edition. Pretoria: Van Schaik Publishers.

Van der Walt, J.L. 2012. The meaning and uses of language test scores: An argument-based approach to validation. *Journal for language teaching*, 46(2): 141-155.

Van der Walt, J.L. & Steyn, F. 2008. The validation of language tests. *Stellenbosch papers in linguistics*, 38: 191-204.

Van der Walt, J.L. & Steyn, H.S. jnr. 2007. Pragmatic validation of a test of academic literacy at tertiary level. *Ensovoort*, 11(2): 138-153.

Van Dyk, T. & Weideman, A. 2004a. Switching constructs: On the selection of an appropriate blueprint for academic literacy assessment. *Journal for language teaching*, 38(1): 1-13.

Van Dyk, T. & Weideman, A. 2004b. Finding the right measure: From blueprint to specification to item type. *Journal for language teaching*, 38(1): 15-24.

Van Dyk, T., Zybrands, H., Cillié, K. & Coetzee, M. 2009. On being reflective practitioners: The evaluation of a writing module for first-year students in the health sciences. *Southern African linguistics and applied language studies*, 27(3): 333-344.

Van Dyk, T.J. 2010. *Konstitutiewe voorwaardes vir die ontwerp van ’n toets van akademiese geletterdheid.* Unpublished doctoral thesis. Bloemfontein: University of the Free State.

Van Dyk, T.J., Van Dyk, L., Blanckenberg, H.C. & Blanckenberg, J. 2007. Van bevreemdende diskoers tot toegangsportaal: E-leer as aanvulling tot ’n akademiese geletterdheidskursus (From alienating discourse to access discourse: E-learning as a supplement to an academic literacy course). *Ensovoort*, 11(2): 154-172.

Van Rooy, B. & Coetzee-Van Rooy, S. 2015. The language issue and academic performance at a South African University. *Southern African linguistics and applied language studies*, 33(1): 31-46. DOI: 10.2989/16073614.2015.1012691.

Van Wyk, A.L. & Greyling, W.J. 2008. Developing reading in a first-year academic literacy course. *Stellenbosch papers in linguistics*, 38: 205-219. DOI: 10.5774/38-0-30.

Van Wyk, P. 2012. SA education standards too low – Oprah. *Beeld*, 16 January 2012. [Online]. Available: http://www.news24.com/SouthAfrica/News/SA-education-standards-too-low-Oprah-20120116 [2014, February 5].

Wallace, C. 2012. Teaching critical reading. In Alsagoff, L., McKay, S., Hu, G. & Renandya, W. (Eds), *Principles and practices for teaching English as an international language*. New York: Routledge. Pp. 282-298.

Webb, V. 2008. Overview of issues at stake. In Lafon, M. & Web, V. (Eds), *The standardization of African languages: Language political realities*. Proceedings of a CENTREPOL workshop held at the University of Pretoria on March 29, 2007, supported by the French Institute of Southern Africa. Pp. 8-21.

Webb, V. 2013. African languages in post-1994 education in South Africa: Our own Titanic? *Southern African linguistics and applied language studies*, 31(2): 173-184.

Weideman, A. 1987. *Applied linguistics as a discipline of design – a foundational study*. Unpublished doctoral thesis. Bloemfontein: University of the Orange Free State.

Weideman, A. 2002. *Designing language teaching: On becoming a reflective professional.* Pretoria: University of Pretoria.

Weideman, A. 2003. Justifying course and task construction: Design considerations for language teaching. *Acta academica* 35(3): 26-48.

Weideman, A. 2006. Assessing academic literacy in a task-based approach. *Language matters*, 37(1): 8-101.

Weideman, A. 2007. The redefinition of applied linguistics: Modernist and postmodernist views. *Southern African linguistics and applied language studies*, 25(4): 589-605.

Weideman, A. 2009a. Constitutive and regulative conditions for the assessment of academic literacy. *Southern African linguistics and applied language studies*, 27(3): 235-251.

Weideman, A. 2009b. *Beyond expression: A systematic study of the foundations of linguistics.* Grand Rapids, Paideia Press.

Weideman, A. 2011. *A framework for the study of linguistics.* Pretoria: Van Schaik.

Weideman, A. 2012. Validation and validity beyond Messick. *Per linguam* 28(2): 1-14.

Weideman, A. 2013. Academic literacy interventions: What are we not yet doing, or not yet doing right? *Journal for language teaching*, 47(2): 11-23.

Weideman, A., 2014. Innovation and reciprocity in applied linguistics. *Literator*, 35(1): 1-10. [Online]. Available: http://dx.doi.org/10.4102/lit. v35i1.1074 [2015, June 6].

Weideman, A. 2017. *Responsible design in applied linguistics: Theory and practice*. Cham, Switzerland: Springer.

Weideman, A., Du Plessis, C. & Steyn, S. (2017 – forthcoming). Diversity, variation and fairness: Equivalence in national level language assessments.

Weideman, A., Tesfamariam, H. & Shaalukeni, L. 2003. Resistance to change in language teaching: Some African case studies. *Southern African linguistics and applied language studies*, 21(1 & 2): 67-76.

Weideman, A. & Van der Slik, F. 2008. The stability of test design: Measuring differences in performance across several administrations of an academic literacy test. *Acta academica*, 40(1): 161-182.

Weigle, S.C. 2002. *Assessing writing.* Cambridge: Cambridge University Press.

Weir, C.J. 2005. *Language testing and validation: An evidence-based approach.* New York: Palgrave Macmillan.

Weir, C. & Milanovic, M. (Eds). 2003. *Studies in language testing 15. Continuity and innovation: Revising the Cambridge proficiency in English examination 1913-2002.* Cambridge: Cambridge University Press.

Weir, C.J., Vidaković, I. & Galaczi, E.D. 2013. *Measured constructs: A history of Cambridge English language examinations 1913-2012. Studies in language testing volume 37.* Cambridge: Cambridge University Press.

Widdowson, H.G. 1984. *Explorations in applied linguistics 2.* Oxford: Oxford University Press.

Wilkins, D. 1976. *Notional syllabuses: A taxonomy and its relevance to foreign language curriculum development.* London: Oxford University Press.

Wilkinson, K. 2014. Africa Check: Why the matric pass rate is not a reliable benchmark of quality education. *Daily Maverick*, 7 January 2014. [Online]. Available: http://www.dailymaverick.co.za/article/2014-01-07-africa-check-why-the-matric-pass-rate-is-not-a-reliable-benchmark-of-quality-education/#.UvIb-fmSyBU] [2014, February 5].

Wyatt-Smith, C. 1997. Teaching and assessing writing: An Australian perspective. *English in education*, 31(3): 8-22.

Xi, X. 2010. How do we go about investigating test fairness? *Language testing*, 27: 147-170.

Yeld, N. 2003. Academic literacy and numeracy profiles: An analysis of some results from the AARP and TELP tests of incoming students (2001/2002 entry years). In *Into higher education – Perspectives on entry thresholds and enrolment systems*. A joint SAUVCA-CTP publication. Pp. 21-52.

Young, R.F. 2012. Social dimensions of language testing. In Fulcher, G. & Davidson, F. (Eds), *The Routledge handbook of language testing.* New York: Routledge. Pp. 178-193.

Yu. G. 2013. The use of summarization tasks: Some lexical and conceptual analyses. *Language assessment quarterly*, 10: 96-109.

Zucker, S. 2003. Fundamentals of standardized testing. Pearson Education, Inc. [Online]. Available: http://www.pearsonassessments.com [2013, June 10].

Abstract

Disproportionately high pass rates and alarming discrepancies in the results of the South African National Senior Certificate (NSC) external examination for Home Language (HL) subjects have undermined the credibility of the state school curriculum and its system of assessment. This has also fuelled allegations that the same standards do not apply to all HL school subjects. Studies commissioned by the Council for Quality Assurance in General and Further Education and Training (Umalusi) to investigate the matter have confirmed that the levels of cognitive challenge do vary across HL examination papers. However, the reasons for this have not been identified and no alternatives have been proposed to improve the situation. In supplementation of the findings of Umalusi, this study seeks to develop a theoretical rationale for greater equivalence of standard in assessment across HLs. It has identified the historical lack of parity of esteem of the HLs and their continued inequitable treatment as academic languages as major factors that have constrained efforts to introduce comparability of standard and construct in the school-leaving examination. Other salient reasons for the varying standards pertain to the problematic design of the HL examination papers, a lack of conceptual clarity of constructs, and a lack of compliance with essential principles for the responsible and fair assessment of language abilities.

In as much as initiatives by the former Department of Education to establish a unified national department and introduce common curricula and standards with centrally set examination papers may be laudable, they have failed to address the weak academic status of the indigenous Bantu HLs beyond the Foundation and Intermediate Phases of schooling. As a result hereof, the pivotal role that these languages could play in fostering epistemological access in other school subjects has been overlooked. The study argues that without comparable treatment of these languages to that of Afrikaans and English, it would be difficult to attain similar standards of teaching and assessment. This is a severe limitation in the quest for equality. It points to the need to raise the academic status of the indigenous Bantu HLs by exploring their use as languages of learning and teaching beyond the initial years of primary school education through bilingual pedagogies and variations of double-medium modes of instruction.

A further recommendation of the study is that the HL examination papers be redesigned to accommodate a common examination component aimed at stimulating the development, inter alia, of academic vocabulary and inferential reasoning. On the basis of the prescribed curriculum, the study has identified the underlying construct of the HL examination to be the assessment of a differentiated language ability in discourse types involving typically different texts, and a generic ability incorporating task-based functional and formal aspects of language. Further to this, learners are required to display a high level of language ability as opposed to the more basic communicative kind needed for other languages offered at First Additional Language (FAL) level. Through the redesign of the exit-level HL examination papers and the introduction of a common examination component that tests generic language ability, a stimulus can be provided for the equitable advancement of the HLs. This would go hand in hand with the development of literacy materials in the HLs to strengthen skills that are needed for further study and workplace environments. At the same time, by adopting an integrated approach to language teaching and testing, the development of the crucial ability to understand, process and produce information in various formats through authentic and relevant tasks can be achieved through language instruction at school.

A third major part of the study focuses on the sophisticated and specialised nature of language assessment as a branch of applied linguistics, and identifies a serious lack of compliance in the HL examination with orthodox principles that support valid and reliable assessment practices. This is particularly problematic in light of the disparities emanating from unjust educational policies in the previous political dispensation. The NSC is in need of socially just and equitable forms of assessment that are ecologically sensitive and reflect political and social accountability on the part of those tasked with overseeing the examination. Examination papers analysed for the purposes of the study show that privileged forms of literacy are still tolerated in some parts of the examination and that unacceptably high ratios of potentially subjective and unreliable scoring are allowed. By means of the proposed theoretical framework for the responsible design of the HL examination papers, a positive washback effect on language teaching is envisaged and a platform created for the eventual attainment of greater equivalence of standard and construct.

Abstrak

Buitengewoon hoë slaagsyfers en sorgwekkende teenstrydighede in die uitslae van die Suid-Afrikaanse Nasionale Seniorsertifikaat (NSS) wat die Huistaal- (HT) eksamens betref, het die geloofwaardigheid van die staatskoolkurrikulum en die meegaande assesseringstelsel ondermyn. Dit het ook aanleiding gegee tot aantygings dat dieselfde standaarde nie op alle Huistaalvakke (HTe) van toepassing is nie. Studies in opdrag van die Raad vir Gehalteversekering in Algemene en Verdere Onderwys en Opleiding (Umalusi) om die saak te ondersoek, het bevestig dat die vlakke van kognitiewe uitdagings tussen HT-vraestelle wissel. Die redes hiervoor of voorgestelde alternatiewe om die situasie te verbeter is egter nog nie geïdentifiseer nie. Ter ondersteuning van Umalusi se bevindings, het die huidige studie ten doel om ʼn teoretiese rasionaal vir groter ekwivalensie by standaarde tussen die huistaalvraestelle te ontwikkel. Dit het die historiese gebrek aan gelykheid van aansien van die HTe, asook die voortgesette ongelyke behandeling daarvan as akademiese tale, geïdentifiseer as belangrike faktore wat pogings kniehalter om ʼn vergelykbare standaard en konstruk in die skooleindeksamen daar te stel. Ander opvallende redes vir die wisselende standaard kan toegeskryf word aan die problematiese ontwerp van die HT-vraestelle, ʼn gebrek aan konseptuele klaarheid wat konstrukte betref, en ʼn gebrek aan voldoening aan die noodsaaklike beginsels by die verantwoordelike en regverdige assessering van taalvermoë.

Inisiatiewe deur die vorige Departement van Onderwys om ʼn verenigde nasionale departement te vestig en gemeenskaplike kurrikulums en standaarde daar te stel met sentraal-opgestelde vraestelle verdien lof. Ondanks hierdie noemenswaardige pogings is die swak akademiese status van die inheemse Bantoe-HTe na die Grondslag- en Intermediêre onderrigfases egter nie onder die loep geneem nie. Die deurslaggewende rol wat hierdie tale kan speel by die bevordering van epistemologiese toegang tot kennis in ander skoolvakke word gevolglik oor die hoof gesien. Die studie redeneer dat, indien hierdie tale nie op gelyke vlak met Afrikaans en Engels hanteer word nie, dit moeilik sal wees om vergelykbare onderrig- en assesseringstandaarde te behaal. Dit is ʼn ernstige beperking in die strewe na gelykheid en dui op die noodsaak daarvan om die akademiese status van die inheemse Bantoe-HTe te verhoog. Een voorstel sou wees om die Bantoe-Hte aan te wend as tale van onderrig en leer vir langer as die aanvangsjare van primêre-

skoolonderrig deur gebruik te maak van tweetalige pedagogiek en ŉ verskeidenheid van dubbelmediumonderrigmetodes.

ŉ Verdere aanbeveling van hierdie studie is dat die HT-eksamenvraestelle herontwerp moet word om ŉ gemeenskaplike eksamenkomponent te akkommodeer, wat onder meer gerig sou wees op die ontwikkeling van akademiese woordeskat en afleibare redenering. Op grond van die voorgeskrewe kurrikulum is die onderliggende konstruk vir die HT-eksamen geïdentifiseer as die assessering van ŉ gedifferensieerde taalvermoë in ŉ aantal diskoerstipes wat tipies verskillende tekste betrek, en ŉ generiese vermoë wat taakgebaseerde funksionele en formele aspekte van taal inkorporeer. Verder word daar van leerders verwag om ŉ hoë vlak van taalvermoë te toon, in teenstelling met die meer basiese kommunikatiewe tipe wat benodig word in die geval van tale wat op Eerste Addisionele Taalvlak aangebied word. Deur middel van die herontwerp van die uittreevlak- HT-eksamenvraestelle en die instel van ŉ gemeenskaplike eksamenkomponent wat generiese taalvermoë toets, kan ŉ stimulus verskaf word vir die gelykwaardige bevordering van die HTe. Dit sal hand aan hand met die ontwikkeling van geletterdheidsmateriaal in die HTe moet gaan, ten einde die vaardighede uit te bou wat vir verdere studie en in die werksomgewing benodig word. Terselfdertyd deur middel van ŉ geïntegreerde benadering tot taalonderrig en -toetsing kan die vermoë om inligting te verstaan, te prosesseer en in verskeie formate voort te bring deur outentieke en relevante take gefasiliteer word.

ŉ Derde belangrike deel van die studie konsentreer op die gesofistikeerde en gespesialiseerde aard van taalassessering binne die veld van toegepaste linguistiek, en identifiseer ŉ ernstige leemte in die HT-eksamen by die voldoening aan ortodokse beginsels wat geldige en betroubare assesseringspraktyke ondersteun. Dit is veral problematies in die lig van die ongelykhede wat voortspruit uit die onbillike onderwysbeleid van die vorige politieke bedeling. Die NSS benodig sosiaal regverdige en billike vorms van assessering wat ekologies sensitief is en wat politieke en sosiale toerekenbaarheid aan die kant van diegene getaak met die ontwerp en administrasie van die eksamenvraestelle weerspieël. Eksamenvraestelle wat vir die doel van hierdie studie ontleed is, toon egter dat bevoorregte vorms van geletterdheid nog steeds voorkom in sommige dele van die eksamen en dat ŉ onaanvaarbaar hoë koers van potensieel subjektiewe en onbetroubare puntetoekenning steeds toegelaat word. By wyse van die

voorgestelde teoretiese raamwerk vir die verantwoordelike ontwerp van die HT-eksamenvraestelle, word ʼn positiewe uitwerking op taalonderrig in die vooruitsig gestel en word ʼn platform geskep vir die uiteindelike bereiking van groter ekwivalensie by beide standaard en konstruk.

Key words

Additive multilingualism

Applied linguistics

Curriculum and Assessment Policy Statement (CAPS)

Constitutive conditions

Construct validity

Content validity

Equivalence in testing

Fairness

Item specifications

Justice

Language-in-Education Policy

Language status

Language teaching

Language testing

National Curriculum Statement (NCS)

Parity of esteem

Regulative conditions

Reliability

Scoring validity

Task types

Test usefulness

Validation

Validity

Washback

Sleutelwoorde

Additiewe veeltaligheid

Betroubaarheid

Ekwivalensie in toetsing

Geldigheid

Geldigheid van puntetoekenning

Gelykheid van aansien

Geregtigheid

Inhoudsgeldigheid

Itemspesifikasies

Konstitutiewe voorwaardes

Konstrukgeldigheid

Kurrikulum- en Assesseringbeleidsverklaring (KABV)

Nasionale Kurrikulumverklaring (NKV)

Regulatiewe voorwaardes

Regverdigheid

Taaktipes

Taal-in-Onderwys-beleid

Taalonderrig

Taalstatus

Taaltoetsing

Toegepaste linguistiek

Toetsbruikbaarheid

Validering

Appendix A: Classification of sub-abilities assessed in English HL Papers 1 and 3 (2008-2012) as listed in the prescribed learning content of CAPS

Students should have an understanding of the following:	
1	abbreviations and acronyms
2	active and passive voice
3	adjectives: comparative, superlative, numerical, demonstrative, relative
4	adverbs: participles of possibility, opinion, time, manner, place, frequency
5	affirmatives and negatives
6	analysing, interpreting graphs
7	analysing, interpreting, evaluating and responding to images, pictures, film scenes, font types and sizes
8	analysing, interpreting, evaluating and responding to language in cartoons/comic strips
9	articles
10	awareness of the socio-political and cultural background of texts and authors
11	cause and effect
12	choice paragraph
13	chronological/sequential order
14	classification paragraph
15	cohesive ties
16	collocations
17	common phrases, proverbs and idiomatic language
18	commonly confused words: homophones, homonyms, homographs, synonyms, antonyms, paronyms
19	compare and contrast
20	concluding paragraph
21	concord
22	definition paragraph

23	denotation and connotation
24	description paragraph
25	direct and indirect speech
26	drawing conclusions
27	evaluating information
28	evaluating how words from various cultural origins have an impact on text, borrowed, inherited and new words (neologisms), etymology
29	evaluation paragraph
30	examining how advertisers get attention
31	examining how layout is a key aspect of popular websites
32	examining how movement and colour play key roles in persuading the reader to move to other sites
33	explanation
34	exposition
35	expressing own point of view/opinion supported by values, beliefs and experiences
36	fact and opinion
37	figurative language and rhetorical devices (simile, metaphor, personification, oxymoron, metonymy, onomatopoeia, hyperbole, contrast, irony, sarcasm, anti-climax, symbolism, euphemism, litotes, paradox, pun, understatement, synecdoche)
38	identify the purpose of including or excluding information
39	identifying images that are sexist, racist, ageist or stereotyped
40	inferring
41	interjections/exclamations/ideophones
42	interrogatives/question forms with modals, positive and negative
43	making connections
44	making notes
45	making predictions
46	making sense of the text

47	noticing the effect of selections and omissions on meaning
48	nouns: abstract, concrete, simple, common, complex, compound, collective, (un)countable, proper, gerunds, predicate and object, gender, plural diminutives, augmentatives, nouns derived from other parts of speech
49	order of importance of information
50	polysemes (multiple-meaning words)
51	prefixes, suffixes and roots of words
52	prepositions/locatives: with a variety of phrasal verbs, movement, place, time
53	procedures
54	pronouns: indefinite, relative clause, demonstrative, possessive, personal, reflexive
55	punctuation: hyphen, colon, semi-colon, apostrophe, quotation marks, parentheses, ellipses
56	reading for main and supporting ideas
57	recognising common allusions
58	recognising emotive and manipulative language, bias, prejudice, discrimination, stereotyping
59	recognising inferences, assumptions and arguments
60	recognising language varieties
61	recognising text types
62	reports
63	reproducing a genre in own writing
64	retelling a story or sentence using different words
65	sentence construction (length and complexity)
66	skimming and scanning text features and book parts
67	spatial order
68	spelling
69	summarising main and supporting ideas
70	synthesising

71	taking into account purpose, audience, topic and genre in writing
72	transition words/conjunctions
73	understanding direct and implied meaning
74	understanding the relationship between language and power
75	understanding how language and images reflect and shape values and attitudes
76	using appropriate words, phrases and expressions in writing
77	using information from other texts to substantiate arguments
78	using one word for a phrase
79	using textual context and cues to determine the meaning of unfamiliar words
80	using reference books to determine the meaning of unfamiliar words
81	verb forms and auxiliaries expressing tense and mood: present, perfect, progressive, passive, future, dual use of some nouns/verbs, gerunds/infinitives, main verbs, transitive, intransitive, finite, non-finite, copulative, regular, irregular, phrasal, stative, verbal extensions (derivatives), auxiliaries linked to modals (subjunctive, imperative, potential, indicative, conditional)
82	visualising
83	word choice and language structures
84	word order
85	writing different kinds of paragraphs (sequential, cause and effect, procedural, comparisons/contrasts, introductory and concluding paragraphs)
86	writing different parts of a paragraph, including introductory, supporting and concluding sentences
87	writing in such a way that there is no ambiguity of meaning, redundancy or inappropriate language
88	writing texts that are coherent using conjunctions and transitional words and phrases
89	writing texts that display own voice with style and register in keeping with the purpose of the writing

Appendix B: Detailed analysis of English HL Paper 1 Section A: Comprehension (2008-2012)

2008														
Question	1.1	1.2.1	1.2.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10.1	1.10.2		12 items
Marks	3	2	2	3	2	3	3	3	3	2	2	2		30 marks
Marking	Sub	Obj	Obj	Sub	Sub	Obj	Sub	Obj	Sub	Obj	Sub	Sub		18 marks subjective
Type	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open		12/12 open items
Sub-ability	35,46	83,25	83,37	83,35	35,46	83,37	26,35	26,27	26,35	83,46	36,44, 47	35,43, 46		6/12 opinion
2009														
Question	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	1.11	1.12		12 items
Marks	2	3	2	1	3	3	3	3	3	2	2	3		30 marks
Marking	Sub	Sub	Sub	Obj	Sub	Sub	Sub	Obj	Sub	Obj	Obj	Sub		22 marks subjective
Type	Open	Open	Open	Closed	Open	Open	Open	Open	Open	Open	Open	Open		11 open items
Sub-ability	35	56	56,69	40,46	46	35	35	46,69	35	23	46,69	35,43		5 opinion
2010														
Question	1.1	1.2	1.3	1.4	1.5.1	1.5.2	1.6	1.7	1.8	1.9	1.10	1.11		12 items
Marks	2	2	2	2	2	3	2	3	2	4	3	3		30 marks
Marking	Obj	Obj	Sub	Sub	Obj	Sub	Obj	Obj	Sub	Sub	Obj	Sub		16 marks subjective
Type	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open		12 open items
Sub-ability	46,69	19,46, 64	40,46	35	17,23	35	46	23	35	19,35	8,40	19,27, 35		5 opinion

2011														
Question	1.1.1	1.1.2	1.2.1	1.2.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	1.11	13 items
Marks	2	2	2	2	2	2	2	2	3	3	2	2	4	30 marks
Marking	Sub	Obj	Obj	Sub	Sub	Sub	Obj	Sub	Sub	Sub	Obj	Sub	Sub	22 marks subjective
Type	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open	13 open items
Sub-ability	26,35	58	46,64	26,35	35,40	10,35, 23, 73	19,26, 27	17,23, 35	26,27, 35	7	83,23	35,51	26,27, 35	8 opinion
2012														
Question	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	1.10	1.11			11 items
Marks	2	3	2	2	3	3	3	3	3	2	4			30 marks
Marking	Sub	Sub	Obj	Sub	Sub	Sub	Sub	Sub	Sub	Sub	Sub			28 marks subjective
Type	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open	Open			11 open items
Sub-ability	46,35	46,35	37	23,35, 46	35	83,17	19,35	26,35	83,35, 37	35	19,35			9 opinion

Appendix C: Detailed analysis of English HL Paper 1 Section B: Summary (2008-2012)

2008		
Question	2.1	2.2
Marks	3	7
Marking	Subjective	Subjective
Specification	44	69,88
2009		
Question	2	
Marks	10	
Marking	Subjective	
Specification	44,69,88	
2010		
Question	2	
Marks	10	
Marking	Subjective	
Specification	44,69	
2011		
Question	2	
Marks	10	
Marking	Subjective	
Specification	44,69	
2012		
Question	2	
Marks	10	
Marking	Subjective	
Specification	44,69	

Appendix D: Detailed analysis of English HL Paper 1 Section C: Language in context (2008-2012)

2008																			
Question	3.1.1	3.1.2	3.2.1	3.2.2	4.1.1	4.1.2	4.2.1	4.2.2	4.2.3	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	18 items
Marks	3	3	2	2	2	2	2	2	2	1	1	1	1	2	1	1	1	1	30 marks
Marking	Sub	Sub	Obj	Sub	Obj	Sub	Sub	Sub	Sub	Obj	Obj	Obj	Obj	Sub	Obj	Obj	Obj	Obj	18 marks subjective
Type	Open	Open	Open	Open	Open	Open	Open	Open	Open	Closed	Closed	Open	Closed	Open	Closed	Open	Closed	Open	13 open-ended items
Sub-ability	7,35	8,35, 83	8,23, 83	7,35, 83	7	35,46	7	23,58 83	35	1	81	79	18	35,55	64	48	21	37	6 opinion
2009																			
Question	3.1.1	3.1.2	3.2	3.3	4.1.1	4.1.2	4.2.1	4.2.2	4.2.3	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	18 items
Marks	2	2	2	4	2	2	2	2	2	1	1	2	1	1	1	1	1	1	30 marks
Marking	Obj	Sub	Sub	Sub	Obj	Sub	Sub	Obj	Sub	Obj	Obj	Obj	Obj	Obj	Obj	Obj	Obj	Obj	14 marks subjective
Type	Open	Open	Open	Open	Open	Open	Open	Open	Open	Closed	Closed	Closed	Closed	Closed	Open	Closed	Open	Closed	11 open-ended items
Sub-ability	46	8,35	7,35	8,35	8	35	7,35	7,8	8,35	4	65	25	68	64	17,83	21	55	55	6 opinion
2010																			
Question	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8.1	5.8.2	18 items
Marks	2	1	3	2	2	2	3	2	3	1	1	2	1	1	1	1	1	1	30 marks
Marking	Obj	Obj	Sub	Sub	Sub	Sub	Obj	Obj	Sub	Obj	Obj	Obj	Obj	Obj	Obj	Obj	Obj	Obj	12 marks subjective
Type	Open	Open	Open	Open	Open	Open	Open	Open	Open	Closed	Closed	Open	Open	Closed	Closed	Closed	Closed	Open	12 open-ended items
Sub-ability	7	7	7,35	7,35	8,35, 83	7,8, 35	7	6,46	6,35	21,54	65,72	65,76	55,68	68	51	55	65	65	5 opinion

2011																			
Question	3.1	3.2	3.3	3.4	4.1.1	4.1.2	4.1.3	4.2.1	4.2.2	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8	5.9	18 items
Marks	2	2	3	3	2	2	2	2	2	1	1	1	1	1	2	1	1	1	30 marks
Marking	Obj	Obj	Sub	Sub	Sub	Obj	Obj	Obj	Sub	Obj	Obj	Obj	Obj	Obj	Obj	Obj	Obj	Obj	10 marks subjective
Type	Open	Open	Open	Open	Open	Open	Open	Open	Open	Closed	Closed	Closed	Closed	Closed	Closed	Closed	Closed	Closed	9 open-ended items
Sub-ability	8,40, 46	8,16	8,35, 48,81	7,35	7,35	8,17, 23	7,8, 37	7,8, 37	7,35	2	83	3	65,84	21	54,65	48	64	65	4 opinion
2012																			
Question	3.1	3.2	3.3	3.4	4.1	4.2	4.3	4.4	5.1	5.2	5.3	5.4	5.5	5.6	5.7	5.8			16 items
Marks	2	2	2	4	2	2	3	3	1	2	1	1	1	1	1	2			30 marks
Marking	Obj	Sub	Sub	Sub	Obj	Obj	Obj	Sub	Obj	Obj	Obj	Obj	Obj	Obj	Obj	Obj			11 marks subjective
Type	Open	Open	Open	Open	Open	Open	Open	Open	Closed	Closed	Closed	Closed	Closed	Closed	Closed	Closed			8 open-ended items
Sub-ability	7	8,35, 83	7,35	8,27, 83	7,8, 83	7	7,8, 37,83	7,35	64,65 ,89	25,65	18	21	48	51	21,65	55			3 opinion

Appendix E: CEFR indicators of language ability (CEFR 2013)

Example of vocabulary descriptors (p. 112):

	VOCABULARY RANGE
C2	*Has a good command of a very broad lexical repertoire including idiomatic expressions and colloquialisms; shows awareness of connotative levels of meaning.*
C1	*Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions; little obvious searching for expressions or avoidance strategies. Good command of idiomatic expressions and colloquialisms.*
B2	*Has a good range of vocabulary for matters connected to his/her field and most general topics. Can vary formulation to avoid frequent repetition, but lexical gaps can still cause hesitation and circumlocution.*
B1	*Has a sufficient vocabulary to express him/herself with some circumlocutions on most topics pertinent to his/her everyday life such as family, hobbies and interests, work, travel, and current events.*
A2	*Has sufficient vocabulary to conduct routine, everyday transactions involving familiar situations and topics.*
	Has a sufficient vocabulary for the expression of basic communicative needs. *Has a sufficient vocabulary for coping with simple survival needs.*
A1	*Has a basic vocabulary repertoire of isolated words and phrases related to particular concrete situations.*

	VOCABULARY CONTROL
C2	*Consistently correct and appropriate use of vocabulary.*
C1	*Occasional minor slips, but no significant vocabulary errors.*
B2	*Lexical accuracy is generally high, though some confusion and incorrect word choice does occur without hindering communication.*
B1	*Shows good control of elementary vocabulary but major errors still occur when expressing more complex thoughts or handling unfamiliar topics and situations.*
A2	*Can control a narrow repertoire dealing with concrete everyday needs.*
A1	*No descriptor available*

Example of grammatical competence, p. 114

	GRAMMATICAL ACCURACY
C2	*Maintains consistent grammatical control of complex language, even while attention is otherwise engaged (e.g. in forward planning, in monitoring others' reactions).*
C1	*Consistently maintains a high degree of grammatical accuracy; errors are rare and difficult to spot.*
B2	*Good grammatical control; occasional 'slips' or non-systematic errors and minor flaws in sentence structure may still occur, but they are rare and can often be corrected in retrospect.*
	Shows a relatively high degree of grammatical control. Does not make mistakes which lead to misunderstanding.
B1	*Communicates with reasonable accuracy in familiar contexts; generally good control though with noticeable mother tongue influence. Errors occur, but it is clear what he/ she is trying to express.*
	Uses reasonably accurately a repertoire of frequently used 'routines' and patterns associated with more predictable situations.
A2	*Uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/ she is trying to say.*
A1	*Shows only limited control of a few simple grammatical structures and sentence patterns in a learnt repertoire.*

Example of sociolinguistic competence, p. 122

	SOCIOLINGUISTIC APPROPRIATENESS
C2	*Has a good command of idiomatic expressions and colloquialisms with awareness of connotative levels of meaning.* *Appreciates fully the sociolinguistic and sociocultural implications of language used by native speakers and can react accordingly.* *Can mediate effectively between speakers of the target language and that of his/her community of origin taking account of sociocultural and sociolinguistic differences.*
C1	*Can recognise a wide range of idiomatic expressions and colloquialisms, appreciating register shifts; may, however, need to confirm occasional details, especially if the accent is unfamiliar.* *Can follow films employing a considerable degree of slang and idiomatic usage.* *Can use language flexibly and effectively for social purposes, including emotional, allusive and joking usage.*
B2	*Can express him or herself confidently, clearly and politely in a formal or informal register, appropriate to the situation and person(s) concerned.*
	Can with some effort keep up with and contribute to group discussions even when speech is fast and colloquial. *Can sustain relationships with native speakers without unintentionally amusing or irritating them or requiring them to behave other than they would with a native speaker.* *Can express him or herself appropriately in situations and avoid crass errors of formulation.*
B1	*Can perform and respond to a wide range of language functions, using their most common exponents in a neutral register.* *Is aware of the salient politeness conventions and acts appropriately.* *Is aware of, and looks out for signs of, the most significant differences between the customs, usages, attitudes, values and beliefs prevalent in the community concerned and those of his or her own.*
A2	*Can perform and respond to basic language functions, such as information exchange and requests and express opinions and attitudes in a simple way.* *Can socialise simply but effectively using the simplest common expressions and following basic routines.*
	Can handle very short social exchanges, using everyday polite forms of greeting and address. Can make and respond to invitations, suggestions, apologies, etc.
A1	*Can establish basic social contact by using the simplest everyday polite forms of: greetings and farewells; introductions; saying please, thank you, sorry, etc.*

Users of the Framework may wish to consider and where appropriate state:

- *what range of greetings, address forms and expletives learners should need/be equipped/be required to a) recognise b) evaluate sociologically c) use themselves;*
- *which politeness conventions learners should need/be equipped/be required to a) recognise and understand b) use themselves;*
- *which forms of impoliteness learners should need/be equipped/be required to a) recognise and understand b) use themselves and in which situations to do so;*
- *which proverbs, clichés and folk idioms learners should need/be equipped/be required to a) recognise and understand b) use themselves;*
- *which registers learners should need/be equipped/be required to a) recognise b) use;*
- *which social groups in the target community and, perhaps, in the international community the learner should need/be equipped/be required to recognise by their use of language.*

Example of knowledge of coherence and cohesion, p. 125

	COHERENCE AND COHESION
C2	*Can create coherent and cohesive text making full and appropriate use of a variety of organisational patterns and a wide range of cohesive devices.*
C1	*Can produce clear, smoothly flowing, well-structured speech, showing controlled use of organisational patterns, connectors and cohesive devices.*
B2	*Can use a variety of linking words efficiently to mark clearly the relationships between ideas.*
	Can use a limited number of cohesive devices to link his/her utterances into clear, coherent discourse, though there may be some 'jumpiness' in a long contribution.
B1	*Can link a series of shorter, discrete simple elements into a connected, linear sequence of points.*
A2	*Can use the most frequently occurring connectors to link simple sentences in order to tell a story or describe something as a simple list of points.*
	Can link groups of words with simple connectors like 'and', 'but' and 'because'.
A1	*Can link words or groups of words with very basic linear connectors like 'and' or 'then'.*