

**Equalising Grade 12 home language
examination papers in a multilingual education
department: The case of Sesotho and English**

JT Mahlasela

 **orcid.org/0000-0002-2200-9167**

Thesis accepted in fulfillment of the requirements for the degree
Doctor of Philosophy in English at the North-West University

Promoter: Prof HG Butler

Co-promoter: Prof TJ van Dyk

Assistant-promoter: Dr H Kotze

Graduation: August 2023

Student number: 11845856

DEDICATION

I dedicate this to my mom, Mrs Paulina Mahlasela. She is the one who has always been pushing that I get this second PhD in the family.

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to the following people who each, in his or her special way, contributed towards the success of this study:

- My promoter, Professor Gustav Butler, who patiently guided me through this study.
- My co-promoter, Professor Tobie Van Dyk, for supplying the much-needed technical knowledge about language testing.
- My assistant promoter, Professor Herculene Kotze, for rendering her profound expertise in translation subjects.
- Professor Albert Weideman, for his profound advice and specialisation about the use of statistical programs.
- Ms Sanet Steyn, for supplying her English data for comparison with the Sesotho data.
- Mr Lebona Mafisa and Mr Lebohang Mathibela for their help with the translation and adaptation of the Tests of Advanced Language Ability (TALA).
- My colleagues from the School of Languages for their support, encouragement and supply of resources.
- Learners and teachers of the four high schools, namely Cedar Secondary School, Nkgopoleng Secondary School, Iketsetseng Secondary School, and Sakhubusha Secondary School (where the experiment was conducted) for the support they gave me.
- The Free State Department of Basic Education for giving me permission to conduct this research at schools under its jurisdiction.
- Significant others for always being there to give a word or two of encouragement.
- My mother and siblings and family-in-law for urging me to push on.
- My wife and kids for being there when I needed them most.
- The Inter-institutional Centre for Language Development and Assessment (ICELDA) for giving permission to use their model of academic testing.

ABSTRACT

The South African Department of Basic Education (DBE) originates from a divided and unequal past. Many educational reforms and transformations took place after the fall of apartheid in 1994. These reforms and transformations were sometimes not adequate and left many unequal circumstances for South African learners.

One of the areas that scholars have expressed uneasiness about, is the results of Grade 12 home language examinations. The main concern seems to be that these results do not reflect the same outcomes for different language groups in this multilingual education department. It was found that Grade 12 home language question papers do not assess the same constructs and are not comparable in terms of reliability and validity. In addition, there was a concern that African home language exam papers were 'easier' than, for example, Afrikaans home language and English home language. The DBE instituted several ministerial committees to address this issue and several interventions were implemented. The Curriculum and Assessment Policy Statement (CAPS) attempted to address this issue by declaring that all 11 official languages of the DBE should be based on the principle of high knowledge and high skill. The Council for Quality Assurance in General and Further Training (Umalusi) commissioned a project in 2013 in which a strategy was sought to entrench this principle. This study is part of the project that sought to address this issue.

One of the strategies which this project proposes is a situation whereby 20% of the total of 300 marks of the Grade 12 home language examination papers should consist of a Test of Advanced Language Ability (TALA). TALA is designed according to the principles of the Inter-institutional Centre for Language Development and Assessment (ICELDA) academic literacy test, called Test of Academic Literacy Levels (TALL).

Key terms: test of advanced language ability, home language, high knowledge and high skill, reliability, validity, construct

KGUTSUFATSO:

Ho netefatsa hore dipampiri tsa kereiti ya 12 tsa puo ya lapeng di a lekana ka hara lefapha la dipuo tse eketsehileng la thuto: taba ya Sesotho le Senyesemane

Lefapha la Thuto ya Motheo (DBE) la Afrika Borwa le simoloha nalaneng e arohaneng le e sa lekaneng. Ho ile ha e ba le dintlafatso le diphetoho tse ngata ka mora hore kgethollo e fele ka 1994. Dintlafatso tsena le diphetoho ka dinako tse ding di ne di sa lekana mme di ile tsa siya maemo a sa lekaneng bakeng sa barutwana ba Afrika Borwa.

E nngwe ya dikarolo tseo barutehi ba hlahisitseng ngongoreho ka tsona, ke sephetho sa dihlahlobo sa Kereiti ya 12 tsa Puo ya lapeng. Ngongoreho e kgolo ho bonahala e le hore sephetho sa dihlopha tse fapaneng sa dipuo sa lefapha lena la dipuo tse eketsehileng ha se bonahatse diphetho tse tshwanang. Ho thotswe hore dipampiri tsa Kereiti ya 12 ha di hlahlobe motheo o tshwanang mme di ke ke tsa bapiswa ka ho tshepahala le ho nepahala ha tsona. Ho ekelletsa moo, ho bile le ngongoreho ya hore dipuo tsa Mafrika tsa lapeng di ne di le 'bonojwana' ho feta Seafrikanse puo ya lapeng le Senyesemane puo ya lapeng. DBE e ile ya theha dikomiti tse mmalwa tsa matona ho sebetsana le taba ena mme dithuso tse mmalwa di ile tsa kenngwa tshebetsong. Setatemente sa Leano la Kharikhulamo le Tekanyetso (SLKT) le ile la leka ho sebetsana le taba ena ka ho ntsha kgweletso ya hore dipuo tsohle tse 11 tsa semmuso tsa DBE di lokela ho thehwa hodima leano la tsebo e hodimo le bokgoni bo hodimo. Lekgotla la Netefaletso ya Boleng ka hara Thupelo e Akaretsang le e Tswelang Pele (Umalusi) le ile la re ho thehwe projeke ka 2013 moo ho lokelang hore ho fumanwe leano la ho hlomamisa leano lena. Thuto ena ke karolo ya projeke ena ya ho fumana tharollo.

O mong wa mekgwa eo projeke ena e e sisinyang, ke boemo boo ho bona 20% ya kakaretso ya matshwao a 300 a Dipampiri tsa hlahlobo ya Kereiti ya 12 ya puo ya lapeng di lokelang ho bopjwa ka Teko ya Bokgoni bo Tswetsweng pele ba Puo (TALA). TALA e qapilwe ka maano a teko ya borutehi ya Inter-institutional Centre for Language Development and Assessment (ICELDA) e bitswang Test of Academic Literacy Levels (TALL).

Mantswe a sehloho: teko ya boleng bo hodimo, puo ya lapeng, tsebo e hodimo le bokgoni bo hodimo, ho tshetjwa, ho nepahala, motheo

Table of contents

DEDICATION	ii
ACKNOWLEDGEMENTS	II
CHAPTER 1: INTRODUCTION.....	1
1.1 Problem statement	3
1.2 Hypothesis	6
1.3 Research questions.....	6
1.4 Research aims and objectives.....	7
1.5 Central theoretical statement	8
1.6 Method of investigation	8
1.6.1 Analysis of the literature	8
1.6.2 Empirical investigation	10
1.6.3 Ethical aspects	11
1.7 Overview of chapters	12
CHAPTER 2: INEQUALITY AND PRINCIPLES OF LANGUAGE TESTING IN GRADE 12 PAPERS.....	14
2.1 Possible causes of inequality in Grade 12 home language papers	14
2.2 Historical causes of inequality between school curricula.....	15
2.2.1 Departmental structures before 1994 and beyond	15
2.2.2 Assessment policies and their impact on inequality	19
2.3 Conclusion.....	29
CHAPTER 3: TRANSLATION IN LANGUAGE TESTING	31

3.1	Introduction	31
3.2	The evolution of translation	31
3.3	Nord's functionalist approach to translation	40
3.4	Gile's sequential model of translation	44
3.5	Vinay and Darbelnet's model	45
3.6	Borrowing	47
3.7	Calque	48
3.8	Literal translation	49
3.8.1	Transposition	50
3.8.2	Modulation	50
3.8.3	Equivalence	51
3.8.4	Adaptation	52
3.8.5	Amplification	53
3.8.6	False friends	54
3.8.7	Loss, gain and compensation	54
3.8.8	Explication	55
3.8.9	Generalisation	55
3.9	Strategies and procedures in test translation specifically	56
3.9.1	Test translation	56
3.9.2	Towards using adaptation as a method of translation	57
3.10	Conclusion	59
CHAPTER 4: VALIDITY AS PROPERTY OF LANGUAGE TESTING		61
4.1	Introduction	61

4.2	The scope of validity in language testing	61
4.2.1	The orthodox perspective	63
4.2.2	The postmodernist perspective	66
4.2.2.1	Descriptive statistics	73
4.2.3	The contemporary perspective	79
4.3	The process of validation and various elements of evidence	86
4.4	Theoretically analysing the usage of the underlying common construct for TALA	89
4.4.1	The reading process applicable to all text types.....	93
4.4.2	Interpretation of graphic and visual texts.....	95
4.4.3	Vocabulary development and language use	96
4.4.4	Sentence structure and the organisation of texts	97
4.4.5	Other components of assessment	97
4.4.6	General underlying construct of HL examination papers.....	98
4.5	Conclusion.....	99
CHAPTER 5: RESEARCH METHODOLOGY	101	
5.1	Introduction	101
5.2	Ethical considerations	103
5.2.1	Acquiring permission from authoritative bodies.....	103
5.2.2	Selection of participants.....	103
5.3	Preparation of the stimulus material	108
5.3.1	Translation of the English TALA into Sesotho.....	109
5.3.1.1	Pre-condition guidelines	113

5.3.1.2	Test development guidelines	115
5.3.1.3	Confirmation guidelines	119
5.3.1.4	Administration guidelines.....	119
5.3.1.5	Score scales and interpretation guidelines.....	120
5.3.1.6	Documentation guidelines	120
5.3.2	Designing the original Sesotho TALA test.....	123
5.4	Administering the test.....	132
5.5	Collection and safe keeping of stimulus material	133
5.6	Methods and measures for TALA validation	133
5.7	Conclusion.....	134
CHAPTER 6: EMPIRICAL RESEARCH RESULTS		135
6.1	Introduction	135
6.2	Data analysis.....	136
6.2.1	Raw scores.....	136
6.2.2	First view of the analysed scores	138
6.2.3	Detailed analysis of the data.....	142
6.2.4	NC-SES-TALA refinement	150
6.2.5	Analysis of the results of the refined NC-SES-TALA.....	173
6.3	Formulating a validity argument for the NC-SES-TALA.....	180
6.4	Conclusion.....	188
CHAPTER 7: REFLECTIONS, CONCLUSIONS AND RECOMMENDATIONS		190
7.1	Introduction	190
7.2	Summary of the main components of the study	191

7.3	Reflection on the translation of the Test of Advanced Language Ability (TALA)	192
7.4	Reflection on the creation of a new Sesotho TALA (NC-SES-TALA)	193
7.5	Reflection on the method of conducting an experiment for equalising English and Sesotho TALAs	193
7.6	Reflection on the results of the Translated Sesotho TALA (T-SES-TALA)	195
7.7	Reflection on the results of Newly Created Sesotho TALA (NC-SES-TALA)	195
7.8	Reflection on the validity argument	196
7.9	Conclusions from the study	197
7.9.1	The level of difficulty of the test and its implications	197
7.9.2	The external and internal validity implications of the study.....	198
7.9.3	Translation versus creating tests anew: what do we learn?	199
7.9.4	Limitations of the study	199
7.9.5	The value of the TALA in the Sesotho HL examination paper	201
7.10	Recommendations.....	202

LIST OF TABLES

Table 2.1: Table illustrating comparison between the old education structure and the new education structure after 1994.....	17
Table 2.2: Specific associations with ministers that took office and key policies on which these phases were based	25
Table 3.1: Examples of borrowing from TALA part 1	47
Table 3.2: Examples of borrowing from TALA part 2	47
Table 3.3: Examples of calque from TALA part 1.....	48
Table 3.4: Examples of calque from TALA part 2.....	48
Table 4.1: Results of the Grade 12 examination of 2011	76
Table 4.2: “Reading tasks requiring critical awareness and advanced language ability” (CAPS, DBE, 2011:23) [adapted from Du Plessis <i>et al.</i> , 2013:13]).....	94
Table 4.3: “Generic abilities for the reading and interpreting of visual texts” (CAPS, DBE, 2012:23 [adapted from Du Plessis <i>et al.</i> , 2013:14]).	96
Table 4.4: “Examples of what learners should understand in terms of language structures and text organisation” (adapted from Du Plessis <i>et al.</i> , 2013:15).....	97
Table 5.1: Numbers of learners who enrolled for the end of year examinations at participating schools.....	105
Table 5.2: Numbers of learners who wrote the tests.....	106
Table 5.3: Divisions of tasks for different question types	112
Table 5.4: Example of items for T-SES-TALA part 1 & 2	122
Table 5.5: Example of items for T-SES-TALA part 1 & 2	122
Table 5.6: “Test components and/or specifications of a test of academic literacy” (Mayburg-Smit, 2017:278).....	126

Table 5.7: Screen shot of items for the NC-SES-TALA part 1 & 2 in alphabetical letters	131
Table 5.8: Screen shot of items for the NC-SES-TALA part 1 & 2 in numbers	132
Table 6.1: Total number of items for T-SES-TALA part 1 & 2 (1st data set).....	137
Table 6.2: Total number of items for NC-SES-TALA part 1 & 2 (2nd data set).....	137
Table 6.3: Key parameters of T-SES-TALA part 1&2.....	139
Table 6.4: Key parameters of NC-SES-TALA part 1&2.....	139
Table 6.5: Key parameters of ENG-TALA.....	139
Table 6.6: T-SES-TALA 187 Item convergence table	146
Table 6.7: NC-SES-TALA 187 Item convergence table	147
Table 6.8: T-SES-TALA subtest intercorrelations	149
Table 6.9: NC-SES-TALA subtest intercorrelations	149
Table 6.10: T-SES-TALA <i>TiaPlus</i> report for the items making up subtest 1 and 2	153
Table 6.11: NC-SES-TALA <i>TiaPlus</i> report for the items making up subtest 1 and 2	153
Table 6.12: T-SES-TALA subtest 1.....	155
Table 6.13: NC-SES-TALA subtest 1.....	156
Table 6.14: NC-SES-TALA <i>TiaPlus</i> report for items making up subtest 3.....	157
Table 6.15: Specifications for subtest 3	159
Table 6.16: NC-SES-TALA subtest 3.....	161
Table 6.17: Specifications for subtest 4	163
Table 6.18: NC-SES-TALA <i>TiaPlus</i> report for items making up subtest 4	163
Table 6.19: NC-SES-TALA subtest 4.....	165
Table 6.20: Specifications for subtest 5	166
Table 6.21: NC-SES-TALA <i>TiaPlus</i> report for items making up subtest 5	167

Table 6.22: NC-SES-TALA subtest 5.....	168
Table 6.23: Specifications for subtest 6	170
Table 6.24: NC-SES-TALA TiaPlus report for items making up subtest 6	170
Table 6.25: NC-SES-TALA subtest 6.....	172
Table 6.26: NC-SES-TALA subtest intercorrelations for refined NC-SES-TALA	177
Table 6.27: Summary statistics of NC-SES-TALA and Ref. NC-SES-TALA.....	178
Table 6.28: “Academic performance of learners in Sesotho HL from 2016–2020” (DBE, 2020:153).....	186
Table 6.29: “Academic performance of learners in English HL from 2016–2020” (DBE, 2020:153).....	186

LIST OF FIGURES

Figure 2.1: “Performance of learners by home language 2011–2013” (DBE, 2014:176)	27
Figure 3.1: “Holmes’ scheme of translation concepts from Pym” (2014:2)	36
Figure 4.1: Diagrammatic representation of the orthodox perspective	63
Figure 4.2: Diagrammatic representation of the postmodernist perspective	66
Figure 4.3: Diagrammatic representation of the contemporary perspective	79
Figure 4.4: Diagrammatic representation of the argument-based approach to validity (a contemporary perspective)	86
Figure 5.1: Screenshot of scrambled text question, ENG-TALA part 1	117
Figure 5.2: Screenshot of multiple-choice question with scale ranging from A–D for ENG- TALA comprehension questions.....	118
Figure 5.3: “The Bachman and Palmer language ability framework” (Van Dyk & Weideman, 2004:8)	129
Figure 6.1: Frequency histogram: T-SES-TALA subgroup 0-Subtest 0.....	140
Figure 6.2: <i>TiaPlus</i> factor analysis for T-SES-TALA part 1&2	140
Figure 6.3: <i>TiaPlus</i> factor analysis of scrambled text for refined NC-SES-TALA.....	174
Figure 6.4: <i>TiaPlus</i> factor analysis of text comprehension for refined NC-SES-TALA (Ref- SES-TALA)	175
Figure 6.5: <i>TiaPlus</i> factor analysis of interpreting graphs and visual information for refined NC-SES-TALA	175
Figure 6.6: <i>TiaPlus</i> factor analysis of vocabulary for refined NC-SES-TALA.....	175
Figure 6.7: <i>TiaPlus</i> factor analysis of grammar and text relations for refined NC-SES- TALA.....	176
Figure 6.8: <i>TiaPlus</i> factor analysis of the complete, refined 60 items for the NC-SES-TALA .	179

Figure 2.1: “Performance of learners by home language 2011–2013” (DBE, 2014:176)	27
Figure 3.1: “Holmes’ scheme of translation concepts from Pym” (2014:2)	36

CHAPTER 1: INTRODUCTION

The South African Department of Basic Education (DBE) originates from a divided and unequal past. Its predecessor, the Department of National Education under the apartheid regime, was made up of several segregated departments before 1994 (Lubisi & Murphy, 2010). There was, for example, the centrally controlled Department of Education and Culture (House of Assembly, HOA) for White people; the Indian and Coloured Department of Education and Culture in the House of Delegates (HOD) and Representatives (HOR) for Indians and Coloureds respectively; and the Department of Education and Training (DET) for black people outside the Bantustans.

Many educational reforms and transformations took place after the fall of apartheid in 1994. All of these reforms and transformations took place from within the new ministry that came to be known as the Department of Education (DoE). It is within this new ministry that numerous curricular and policy changes took effect. All of these changes occurred according to each minister that took office from 1994 to the present (Kanjee & Sayed, 2013: 442-469). These changes, although meant to address the inequality and unfairness that were left by the apartheid system, did not adequately address this problematic situation (Jansen, 1998;1999). For this reason, several ministerial task teams were appointed to investigate and recommend solutions (DBE, 2014; Ministerial Committee, 1998). Among the many elements of inequality and unfairness that were identified by these task teams, the issue of language inequality featured prominently. It was discovered, for example, that the language syllabi of the African languages did not promote higher cognitive thinking, they repeated the same material every year and there was a lack of a common framework of assessment between home languages (Ministerial committee, 1998:4; Umalusi, 2005). The latest task team report further revealed that certain National Senior Certificate (NSC) home language (HL) examination marks were poor predictors of whether or not the learners were likely to pass the NSC overall (DBE, 2014:178). The reason for this is that, although some learners gained high marks in the end of year Grade 12 African HL examinations, they still failed their NSC examinations which was not the case for English and Afrikaans HLs (DBE, 2014:178).

It is therefore clear that despite all the efforts towards equalising education in South Africa after apartheid, inequality still persists with specific reference to Grade 12 HL examination papers. This inequality stems from a combination of factors emanating from the legacy of apartheid and inadequate attempts to transform the South African education system by the DoE. These factors emanating from the legacy of apartheid and concomitant inadequate attempts by the new government to transform the education system will be expanded upon in chapter two.

In order to address this inequality, a wider study – of which this study forms part – was commissioned in 2013 to propose an appropriate strategy for contributing towards the equalisation of Grade 12 HL examination papers. This research project was commissioned by the Council for Quality Assurance in General and Further Training (Umalusi) for the equalisation of the South African NSC HL examination papers. It was undertaken jointly between the Inter-institutional Centre for Language Development and Assessment (ICELDA) which, at the time, was a formal partnership between four multilingual South African universities (University of Pretoria, Stellenbosch University, North-West University and the University of the Free State) and Umalusi. Although the wider project is aimed at including all eleven official HLs in the end, it has, to date only focused on the following HLs: Afrikaans, English, Sesotho (to which this study belongs) and Sepedi.

In order to create equivalence between the Grade 12 HL examination papers in the broader study, a strategy has been accepted where, for all involved languages, a common construct towards assessment would be used (Du Plessis *et al.*, 2014). This common construct may eventually contribute to the development of a comprehensive, common framework of assessment for home languages, and by implication, also towards a comprehensive equalisation of Grade 12 HL examination papers for all official South African languages. The proposal in the bigger project is, therefore, that this common construct be used in order to develop a shared 20% of the 300 marks that comprise the exit-level HL papers. It is further suggested that the shared 20% is generally referred to as the Test of Advanced Language Ability (TALA) (Du Plessis *et al.*, 2016).

1.1 Problem statement

As a result, the main problem investigated by this study is what the most appropriate strategy would be for creating equivalence between the different TALAs that are being developed for the official South African languages with specific reference to Sesotho and English. Two strategies that could be employed towards creating such equivalence in the shared TALAs are, firstly, that such a section for Sesotho could be developed anew based on the shared or common construct. The second strategy is to translate the already developed English TALA (Steyn, 2013) into Sesotho.

In order to establish which one of the two strategies is the most adequate and appropriate, two different types of tests, an original (new) Sesotho TALA and a translated (from the English version) Sesotho TALA, would have to be developed/translated and piloted. 'Adequacy' and 'appropriacy' in this context refer to the degree of equivalence that could be established between each of the two types of tests referred to above and the English TALA. An important focus of building an argument about test equivalence is that each of the two tests will have to be investigated for reliability and validity. Perspectives about creating equivalent tests with regard to the concepts of validity and reliability as espoused by prominent researchers in this field (among others, Bachman, 1990; Green, 2014; Messick, 1989; Van Dyk, 2011; Weideman, 2009; Weir, 2005a) will be explored.

The position which validity occupies in this study as one of the basic criteria for language testing is important in this exploration. Validity can be understood from an orthodox as well as a more contemporary perspective on language testing (Du Plessis, 2017). The orthodox conception of validity is that it is "an inherent characteristic of a language test and that it exists as a real construct in the minds of test-takers" (Van der Walt & Steyn, 2008:192). However, the prevailing view thereof is that "it is not a characteristic of a test but a property of the interpretation of test scores". This latter conception was derived from the perspective of the American psychologist, Samuel Messick (1990). This perspective led to the notion of a unified model of validity. This was brought about by the conceptualisation of validity that portrayed construct validity

as “a central underlying component with content and criterion validity as aspects of construct validity” (Messick, 1989:20). Based on this model, the same underlying construct is a basic requirement towards the creation of equivalent tests in language assessment. For this study, a common underlying construct for all home languages within the South African DBE was established when the original English TALA was designed as the anchor study for the wider project (Du Plessis *et al.*, 2016). Messick’s perspective on validity also comprises a four-way classification of validity which includes considerations of evidence and consequence. As a result, the consequential validity that emanated from these considerations came to be based on the social consequence that a test should have on individuals and the evidence about the judgement that should emanate from empirical evidence of the scores (Van der Walt & Steyn, 2008).

The unitary concept of validity advanced by Messick was also adopted by Bachman (1990) and came to be widely accepted. Another prominent researcher who adopted Messick’s model was Weir (2005a :14). He came to accept validity as a superordinate category that included other smaller categories such as ‘context validity, theory-based validity, scoring validity and external validity’. Shaw and Weir (2007) identify these categories as building blocks of construct validity.

Scoring validity as identified by these researchers therefore assimilates reliability into construct validity. To describe this relationship, Weir (2005a:14), maintains that reliability is an aspect of validity. Based on this, one can understand why Van der Walt and Steyn (2008:194) explain the process of validation as containing two stages: “(i) where the test produces the results and the evidence of the candidate’s ability; and (ii) the empirical validation of the procedures by which the test judgements are reached”. In these stages, evidence that may be collected relates to construct validity, content validity, criterion validity and reliability indexes (Van der Walt & Steyn, 2008:194).

This study therefore, includes various components of Messick’s validity model in the construction of a validity argument to show which of the two tests, i.e. the Translated Sesotho TALA and the Newly Created Sesotho TALA, is the most adequate and

appropriate test to be used as part of the Sesotho Grade 12 HL final examination. It amongst others, investigates reliability as part of this unitary whole that makes up the argument. The validity arguments for the two tests would then be compared to determine the most suitable test. With regard to reliability, this study is seeking to explore the view of prominent research about the impact of reliability on equivalence in language testing. Tests in this study are meant to be high stakes in that they are meant to be part of the Grade 12 exit-level examination. For this reason, and reliability purposes, the norm for a Cronbach's alpha value is generally expected to be 0,8 or higher (Weir, 2005a). Differences in terms of the tests' reliability alphas are being determined in line with the prevailing practice in language assessment to ensure that test scores are free from errors and can be relied upon for decision making. The internal consistency of each test as a whole is determined as well as for the different sections of the individual test. In order to determine the reliability difference between the tests, a correlation coefficient for each test is calculated using *TiaPlus*. Results obtained from *TiaPlus* for the Translated Sesotho TALA and Newly Created Sesotho TALA is to determine which of the two tests is better in terms of reliability for the whole test, reliability between subtests and the whole test, as well as reliability regarding an intercorrelation between different subtests. An additional statistical programme, *Winsteps*, is used to verify results that are obtained from *TiaPlus*. How this was done, is explained in Chapters 5 and 6.

For the second strategy, which is the translation of the already developed English TALA, there was a need to follow an appropriate translation method to ensure translation equivalence between the English TALA and the Translated Sesotho TALA. In order to do this, Nord's Functionalist approach was employed. What makes this approach appropriate for the purposes of this study is that it uses the contextual purpose/function of the text as the key concept driving the main objective for translation (Butler, 2017). This principle is particularly important for the current study since the translation must retain the function of the test as an academic literacy test – the purpose for which it is going to be used is therefore the main concern in the translation. The Sesotho version must therefore test the same academic literacy

abilities on the same level as the English TALA. An attempt was made to effect a literal translation of the source text (ST) first, and where a loss of meaning was experienced because of certain contextual factors, other translation procedures and strategies to achieve the objective of the translation according to the basic aim of the functionalist approach (Van Dyk *et al.*, 2011) was employed. Such strategies that may be employed could include naturalisation (if a word or phrase that allows for this strategy to be employed is available in the target language) and paraphrasing, as well as procedures such as cultural equivalence, functional equivalence, and descriptive equivalence (Newmark, 1998:81-93). The strategy of back translation was further used to ensure that similar understanding, both in the source text (ST) and the translated target text (TT), was achieved. Employing a strategy of back translation normally reveals discrepancies in meaning between the ST and the TT. Based on this problem statement, the hypothesis for the study is presented below.

1.2 Hypothesis

The hypothesis for this study is that, if one of the two TALAs, i.e. the Translated Sesotho TALA and the Newly Created Sesotho TALA, shows better internal and external validity than the other, the TALA with a better internal and external validity shall be considered the more adequate TALA. Consequently, the more adequate TALA shall be recommended to form 20% of the 300 marks for the Grade 12 home language examination paper because of its better validity.

1.3 Research questions

Based on the problem statement above, the primary research question in this study is therefore:

Which test between the Translated Sesotho TALA and the Newly Created Sesotho TALA shows a higher level of equivalence with the English TALA when investigated empirically?

In order to answer the primary research question above, the following secondary questions were answered:

1. When considering both the Translated Sesotho TALA and the Newly Created Sesotho TALA, which of these two TALAs shows better evidence of internal validity with the English TALA?
2. When considering both the Translated Sesotho TALA and the Newly Created Sesotho TALA, which of these two TALAs shows better evidence of external validity with the English TALA?

1.4 Research aims and objectives

As suggested by the topic of this study, the general aim of the study is to find the most adequate strategy that can be used to contribute to the equalisation of Grade 12 HL examination papers for Sesotho and English. In order to achieve this aim with specific reference to the above-mentioned two languages, empirical research with a primary focus on the equalisation of three tests, viz. the English TALA, Translated Sesotho TALA and Newly Created Sesotho TALA was conducted. Additionally, this research aims to provide support for the argument of the broader study that Grade 12 HL examination papers can be equalised, provided proper scientific methods of equalisation are applied. With regard to a determination of the adequacy of the two Sesotho TALAs in terms of their potential use in Grade 12 HL examination papers, statistical evidence related to test validation was presented that would provide evidence for an argument about such adequacy. In the end, the measure of adequacy therefore will be related to the degree to which the tests show full scale invariance (equivalence) and whether they are free from any type of bias (including construct, method and item bias) (Koch, 2009:306-312).

As already mentioned, the study would therefore include a description of the development of the new Sesotho TALA and the translation of the already existing English TALA. After these two tests (the newly developed as well as the translated one) were piloted, their effectiveness was statistically correlated to evaluate which is

the most adequate between the two in terms of their level of equivalence with the English TALA.

1.5 Central theoretical statement

The central theoretical statement of this study is based on the notion that if 20% of the 300 marks of the language paper of the Grade 12 HL examination (in this case, that of the Sesotho and English papers) can be made up of either the original Sesotho TALA or the translated version thereof (depending on which of the two shows the highest level of equivalence with the English TALA), then a higher level of equality between the Sesotho HL paper and the English HL paper can be achieved. According to Du Plessis *et al.* (2016:16), this would create “a degree of overlap and equivalence of a reliable component of assessment which makes it possible to make statistical adjustments across the total that would further aid comparability and possible mark adjustment in order to ensure fairness”.

1.6 Method of investigation

1.6.1 Analysis of the literature

Causes of inequality: A comprehensive literature survey was conducted into the causes of the unequal state of affairs in Grade 12 HL examination papers both from its past history and its present policies. Part of this information is reflected in the background to this study above. Secondly, this information was used to identify gaps in the present education system that led to these inequalities and how these gaps can be filled.

Validity as property of language testing: Information about the conceptualisation of the concept of validity from the orthodox perspective to the modern perspective was explored. This information would help to situate the concept of validity in the current perspective. The current perspective is one in which validity is treated as the property of the interpretation of test scores and not as an attribute of the test (Van der Walt & Steyn, 2008:192). This perspective has been re-worked by different researchers over

time to the point where reliability is currently considered to be an aspect of validity. Determining the reliability of test items as well as the overall test scores is still considered standard practice in test assessment and would be included in this study as part of the internal validation of the tests. Validation of other test factors such as the construct, test administration and test-takers would be used for the process of operationalisation of validity.

The issue of construct validity is key in this study because construct equivalence has been found to be lacking across home language papers examination in the South African DBE (Weideman, 2016:7). As Taylor and Weir (2008:130-143) emphasise, there is a need to design examinations that are based on a common construct for different languages if one needs to equalise them. The reason for this is that, if the construct is common, the ability or trait to be measured will also be common (Patterson & Weideman, 2013). For this reason, the anchor study identified a common construct when the English TALA was created at the beginning of the research project (Steyn, 2013). The common construct for HLs that was adopted by the project team is the following:

The assessment of 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 *et al.*, 2016).

The above-defined common underlying construct was used accordingly in this study when the new Sesotho TALA was created. Because the English TALA was based on this construct, the Translated Sesotho TALA was based on the same construct as well.

Reliability as element of language testing: Since the notion of reliability plays a prominent role in language testing and equivalence, relevant literature about the importance of reliability in language testing, and, more specifically, the establishment of equivalence in testing, will be discussed. A comprehensive literature survey will be completed about how to determine the reliability of a test. After the reliability measures (alpha values) of each test have been determined, the reliability indices of these will

be compared as part of an argument towards determining which test performs better than the other with regard to this measure.

Translation approaches towards creating equivalence between the Sesotho and English TALAs: Literature outlining which approach and method would be most suitable for translating a test from one language to the other would be surveyed in order to guide the activity of translation.

Translation approaches would be reviewed in order to situate Nord's functionalist approach within the broader discussion on translation studies. A functionalist approach to translation is ideal for the translation of tests because it incorporates translation strategies, such as adaptation (Van Dyk *et al.*, 2011), which allows the translator the flexibility to adapt a test where a literal translation does not lead to the adequate transference of meaning to a target text.

The literature on back translation as a method of ensuring translation accuracy will also be discussed. Back translation is considered appropriate for this translation because, as stated by Van Dyk *et al.* (2011), it is a method that is widely used and has been internationally accepted for the translation of tests. Moreover, it is commonly considered that the validation of translation hinges upon the process of double translation, both forward and back.

1.6.2 Empirical investigation

Preparation of stimulus material, data collection and population: As noted above, the English TALA has already been developed. Data with regard to the reliability of this test will, therefore, be accessed from the larger project's data base.

The original English TALA will be translated into Sesotho. The Translated Sesotho TALA will then be translated back into English to ensure similarity of understanding and meaning between the ST and the TT. Secondly, the new Sesotho TALA will be developed in line with language testing requirements of reliability and validity of which, most important, the shared test construct will be used.

Participants: The Translated Sesotho TALA together with the Newly Created Sesotho TALA will be administered to Sesotho HL learners of several secondary schools in one of the Northern Free State townships, called Zamdela. The same participants will write both TALAs. This township has four secondary schools which all offer Sesotho HL at Grade 12 level.

Data analyses: The data collected from the administration of the two Sesotho TALAs will be subjected to item analysis by using the *TiaPlus* and *Winsteps* statistical programmes to determine their internal and external validity. This is necessary as the internal and external consistency reflect the reliability or empirical properties of each test (Van Dyk *et al.*, 2011:162; Vander Slik & Weideman, 2008:165). *TiaPlus* uses both CTT analysis and Rasch analysis to produce parameters for both internal and external validities. The reliability measure which is used for the *TiaPlus* statistical package is Cronbach's α (Van der Slik & Weideman, 2008:165). This will inform the researcher about the adequacy of the two tests, and, more specifically, whether one test proves to be more adequate than the other. Furthermore, both the Translated Sesotho TALA and the Newly Created Sesotho TALA will be compared with the English TALA in order to evaluate which one between the Translated Sesotho TALA and the Newly Created Sesotho TALA is more similar to the English TALA. In the end, the test with the highest alpha value between the Translated Sesotho TALA and the Newly Created Sesotho TALA will be the one which is more adequate as compared to the English TALA.

1.6.3 Ethical aspects

The researcher applied to the North-West University for ethics clearance before any research activities with school learners as participants in this study were conducted. There was also an application to the Free State Department of Education before any data collection or experimental research was done at schools in its jurisdiction. Once permission was granted to conduct research at school sites by the Provincial Department of Education, application was also made to schools for permission to conduct research at these centres (evidence of this is included in appendix B2).

Learners were issued with consent forms on which it was made clear that participation in the study was voluntary and that they could withdraw from the study willingly if they so wished. Consent letters were also sent to parents to give permission for their children to take part in this study.

1.7 Overview of chapters

Chapters in this study are arranged according to the following format.

Chapter 1 provides the background against which this study is situated. It makes specific reference to how it came about that Grade 12 HL examination papers in the South African Department of Basic Education are not considered equal in terms of how they assess the proficiency and achievement of learners. It also presents the problem statement, hypothesis, research questions, research aims and objectives, central theoretical statement and method of investigation for the study.

Chapter 2 provides a comprehensive discussion of the literature with regard the inequality in home language examination papers in the DBE as well as principles of language testing. These include the historical causes of inequality among home language Grade 12 examination question papers. Historical causes of inequality are outlined from the time of apartheid to the time of democracy in South Africa. This discussion also includes assessment policies of this department and their impact on inequality.

Chapter 3 deals with relevant literature concerning translation in language testing. It includes a discussion about how translation developed from a mere practice to a fully fledged science. It further focuses on different theories of translation and their history, as well as models of translation, translation strategies and methods.

The literature about validity and reliability in language testing is reviewed in Chapter 4. The chapter presents a critical discussion of the scope of validity in language testing with a focus on the orthodox, postmodernist, contemporary and argument-based perspectives. It further presents an argument about the importance of making use of

12

a common test construct for the development of TALAs in the different official South African languages.

Chapter 5 presents the research methodology that has been employed for the study. It explains in detail what research methods were utilised for data collection, including issues such as the preparation of the stimulus material, administering the test, collection and safe-keeping of material, methods and measures for the validation of TALA and how ethical considerations were addressed.

Chapter 6 deals with the results of the research. It presents an analysis of the data that includes the overall results of the refinement of the TALA as well as the refinement of the newly created TALA specifically. It concludes with the formulation of a validity argument for one of the two tests between the Translated Sesotho TALA and the Newly Created Sesotho TALA.

This chapter reflects on the most salient findings of the study. It focuses on prominent issues regarding the translation of the ENG TALA into Sesotho as well as the creation of a new Sesotho TALA. It also reflects on the method that was used for conducting the experiment towards the equalisation of HL examination papers, the results of the data that was collected through administering the translated TALA and the newly created TALA, and the validity argument. It concludes by providing recommendations for further research that have been identified in the study.

CHAPTER 2: INEQUALITY AND PRINCIPLES OF LANGUAGE TESTING IN GRADE 12 PAPERS

2.1 Possible causes of inequality in Grade 12 home language papers

The home language curriculum in the South African primary and secondary school system and the changes implemented to the curriculum over an extended period of time form the basis from which this study originates. There has always been a concern about the manner in which the curriculum of the South African education system for schools as a whole was treated from the time of apartheid until now. The term often used for this, is the “divided curriculum”. The “divided curriculum” applies to the time when the home language curriculum was viewed as discriminatory and biased, while during the time of democracy it was blamed for its defective reforms and transformations (Jansen, 1998, 1999).

The literature survey that follows puts into perspective the impact of curriculum policies of the apartheid era as well as the curriculum policies of the post-apartheid era on the curriculum with regard to the inequality that exists between home language examination papers that we experience today. The chapter focuses firstly on a description of educational structures before 1994, as well as such structures post-1994. Curriculum aspects include issues of curriculum systems, curriculum statements, assessment issues, and teaching approaches. This chapter also illustrates how education politics caused inequality within different school subjects, including home languages that were offered to different population groups. In addition to the historical causes of inequality in how home languages were treated, this chapter explores literature dealing with strategies to address issues of equivalence between languages.

2.2 Historical causes of inequality between school curricula

2.2.1 Departmental structures before 1994 and beyond

This section provides a historical overview of various structures of the South African education system before 1994 to illustrate the conceptual basis from which the inequality between different home languages originates. Historically, the South African education system has always been divided in terms of the curriculum that it offered to different education departments. The Department of Education and Culture in the House of Assembly (HOA) for white people under the apartheid regime also encompassed sections of the House of Delegates (HOD) for the Indians and sections of the House of Representatives (HOR) for Coloureds. However, the Department of Education and Culture for white people was more privileged as can be seen later on, than those of the HOD and HOR as well as for those of the Department of Education and Training (DET) for black people outside the Bantustans (Lubisi & Murphy, 2010). It is necessary at this point to also highlight that all these departments were controlled centrally by the apartheid Department of National Education (DNE).

The manner in which the apartheid state made sure that education was centrally controlled was by a system which required that the curriculum be developed from the DNE. This was done through the development of the curriculum by white provincial departments which in turn fed the white HOA department at national level. The white provincial departments were led by the so-called Committee of Heads of Education (CHE) which belonged to white officials only. This arrangement fuelled the notion that the DNE's curriculum's intention was aimed at preparing "black learners for servitude, while the white pupils were to be trained for overlordship" (Lubisi & Murphy, 2010:255).

The school curriculum before 1994 was applied to a school system that was divided into a model of private and state schools. But, just at the time when the democratic climate intensified in 1990, all the HOA schools were offered an option to change to a four-class model or remain the way they were. This was because most of the HOA schools were underutilised and pupils of colour were afforded an opportunity to join

these schools if they so wanted. Underutilised schools, which were not prepared to open for pupils of colour, could face closure unless they chose any of the models put on the table. The first model was for the school to retain the status which it possessed or, by a majority vote, choose any of the three models, i.e. model A (to be converted into a private school), model B (to be converted into a state school where parents could determine their own admission policy and where a 60% majority of pupils would remain white) or model C (to be converted into a state-aided school or 'semi-private school under the control of the school Governing Body' (SGB)).

The state department of education for black pupils outside of the Bantustans was called the Department of Education and Training (DET) and was run by the Department of Education (DoE) which became the Department of National Education (DNE) in 1992 for black people. In addition, this department was rooted into the ideology of Bantu education which was based on providing separate education for different race groups (Umalusi, 2007:6).

Earlier, the private schools were very well resourced and financially strong. They were owned by churches and later by profit motivated companies. However, state schools were functioning on an unequal basis with regard to the population group for which they were intended. State schools in towns were meant for the white population and were also well-resourced and well-funded, while state schools in townships were meant for the black population and were poorly resourced. State schools in towns were opened for black learners who could afford to travel to towns during the early 1990s. There were two models of state schooling on farms. They operated more or less like township schools and town schools. There were farm schools for black learners and farm schools for white learners. State farm schools for black learners were poorly equipped and lacked resources while state farm schools for white learners were well-equipped. State farm schools for black learners were managed by the owners of the farms on which they were situated. They were meant for children of the farm tenants who worked on these farms, while state farm schools for white learners were meant

for the children of farm owners of adjoining farms and managed by one member of the farm community from the adjoining farms.

All schools, both state and private, were divided initially into two levels, viz. primary and secondary school levels. The primary school was divided into lower primary school for the two sub-standards A and B (SSA and SSB) and standards 1 and 2; and senior primary school for standards 3 to 5 (Lubisi & Murphy, 2010:258). The secondary school was in turn divided into standards 6 to 10 of which standard 6 and 7 were junior secondary and standards 8–10 were senior secondary (Lubisi & Murphy, 2010:258). This structure was changed into the General Education and Training (GET) band which was divided into Grade R (a year of reception and orientation); Foundation Phase (Grade 1–3); Intermediate Phase (Grade 4–6); Senior Phase (Grade 7–9); and Further Education and Training (FET) which was divided into Grades 10–12 later on during the time of democracy after 1994. These divisions can be diagrammatically divided as follows to illustrate a comparison between the old structure and the new structure:

Table 2.1: Table illustrating comparison between the old education structure and the new education structure after 1994

	PRIMARY SCHOOLING				SECONDARY SCHOOLING	
Old structure	Lower primary SSA–STD 2		Senior Primary STD 3–STD 5		Junior Secondary STD 6–STD 8	Senior Secondary STD 9–STD 10
	GET BAND				FET BAND	
New Structure	Grade R	Foundation Phase Grades 1-3	Intermediate Phase Grades 4-6	Senior Phase Grades 7-9	Grades 10–12	

The structure of the education system after democracy changed dramatically. Already from 1990, many changes began taking place. This was caused by the unbanning of liberation movements such as the African National Congress (ANC) and the Pan Africanist Congress (PAC) and the release of Nelson Mandela from prison by the then apartheid regime. Subsequently, the Department of National Education was renamed the Department of Education (DoE). It was within this department where a number of crucial reforms and transformations took place. As indicated in Chapter 1 of this study, these reforms and transformations took place according to each new minister that took office (Kanjee & Sayed, 2013:447). These ministers each brought changes to the curriculum and that added to the instability of the curriculum. Each tried to implement

changes that they thought would address the criticism that was levelled against the preceding minister.

Notable in the changes that took place after 1994, was the introduction of the National Qualifications Framework (NQF). This body was formed to “create an integrated national framework for learning achievements, facilitate access to, and mobility and progression within the education and training career paths...” (DoE, 1995:1). For this reason, the NQF was entrusted with the responsibility to set standards for education in South Africa. Politically, the NQF served the purpose of equalising the standards between different school systems, namely, the former Model C schools, Department of Education and Training schools and private schools. In addition, according to Kanjee and Sayed (2013:447), “the NQF served as the regulatory mechanism for the implementation of an outcomes-based curriculum”.

Shortly after the South African Qualification Bill (DoE, 1997:1) was passed for the NQF, this bill adopted the outcomes-based approach (outcomes-based education – OBE) to curriculum reform. This teaching and learning approach formed part of the Curriculum 2005 framework. It is within this framework that the schooling system was divided into the broad GET and the FET sections that were mentioned earlier. Central to the issue of the NQF and the outcomes-based teaching and learning approach was that the approach to assessment differed from previous assessment approaches.

2.2.2 Assessment policies and their impact on inequality

Assessment is a key aspect of the curriculum and is, therefore, a relevant area to indicate the extent to which the curriculum between the different racial and ethnic groups differed. As the major focus of this study, assessment is the one element that clearly indicates the extent of the inequality between different subjects for different population groups. Pupils of different education departments were subjected to the end-of-year summative assessment which used to be known as the final examinations, during the time of apartheid. A certificate was issued for the Standard 5, Standard 8 and finally for Standard 10 (also known as Matric) learners before 1994.

Before 1994, there were different examinations for different racial and ethnic groups (i.e., different groups wrote different examination papers). Even examination bodies were different – there were different examination bodies for different provinces, including the Transvaal Education Department (TED), the Cape Education Department (CED), the Natal Education Department (NED), and the Orange Free State Education Department (OFSED). Some of these departments were granted leave to run their own matriculation examination as time went on. This permission to run their own matriculation examination was the prerogative of the Joint Matriculation Board (JMB). Matric, the final ‘standard’ of school, has always been considered the most important standard because it symbolised the end of school and the beginning of work or tertiary education for the youth (National Education Policy Initiative, 1992:25).

Although every department was responsible for setting its own standard ten papers, for issues of moderation, the JMB was responsible for the quality control of all standard ten examination papers. Since papers used to be set at provincial level, this kind of control was done from that level upwards. That meant that no paper could go through the system without due scrutiny for compliance with the apartheid discriminatory laws, even though perhaps primarily meant for quality assurance and equality. The control by the JMB was strict and robust, something which the present system could have learnt from. Lubisi and Murphy (2010:260) summarise this control process as follows:

Firstly, all examination papers set by the state departments of education had to be scrutinised and the marking of scripts had to be checked by JMB subject moderators. Secondly, the statistics of matric exam results in the state departments were also scrutinised. Thirdly, the JMB had the power to withdraw the right of any department to run the matric exam; a right which was annually renewable through a formal application. Fourthly, the JMB exercised control over school syllabi.

Compliance with the apartheid discriminatory laws was unfortunately part of this strict and robust control measures. This is obvious if we consider that different subject syllabi were racially and ethnically different (Jansen, 2010:59). With regard to matric results, which is the focal point of this study, for instance, Lubisi and Murphy

(2010:262) indicate that “the African matric results were manipulated by the Apartheid government so as to suit the socio-economic and political development of South Africa”. Moreover, Trumpelmann (1991) asserts that because of the logistical difficulty ‘to conduct oral exams’, the JMB was forced to allow an oral component in the examination only for “official” languages (English and Afrikaans) where they were offered as second languages. Lubisi and Murphy (2010) further mention some of the factors that might have led to the inequality between the same subjects that were taught at different departmental schools in South Africa. These factors will be discussed comprehensively later on when issues of assessment are discussed.

The JMB as the quality control body came to be replaced by Umalusi – the Council for Quality Assurance in General and Further Training – in the new dispensation. It is within this new quality control body that a number of changes took place in an attempt to make the education system more equitable and fair for all citizens.

In addition to the ordinary summative assessment that included quarterly tests and examinations at the end of the year as the most important milestone for learners before democracy, the issue of year marks as part of assessment had been part of a debate in education that started as early as the 1960s. It was first conceived as the Continuous Assessment (CA) debate in sub-Saharan Africa. There was a strong argument for the implementation of CA but there were strong arguments against it as well (Lubisi & Murphy, 2010). Some of the arguments for its implementation were based on the notion that it could reduce pressure for teachers and pupils and that it could reduce the amount of writing for the standard ten evaluation. However, arguments against CA, based on practical observations, indicated that, in fact, work load for both teachers and pupils increased and that it required considerable capacity from teachers (Lubisi & Murphy, 2010). In South Africa, CA was first introduced by the Natal Education Department (NED) in 1982. It was also advocated strongly by the now defunct Transvaal Education Department (TED), but was never officially used widely in South Africa during that period. As can be seen from this history, all these departments, which had a short spell with CA, were whites only departments of education.

The form of continuous assessment as we know it today was introduced with Curriculum 2005 in the new dispensation and was referred to according to its new acronym, CASS (referred to as CA in the old dispensation). Curriculum 2005 introduced a new approach to teaching that came to be known as Outcomes Based Education (OBE). Outcomes Based Education (OBE), according to Sedibe (2006: 277), meant that each subject had to include critical outcomes and specific outcomes. Subjects were referred to as learning areas after the implementation of Curriculum 2005. Critical outcomes were inclusive of the knowledge, values and skills for democratic citizenry, while specific outcomes were comprised of what learners were meant to be able to do at the end of the lesson. For this reason, Curriculum 2005 became known as an OBE-curriculum. It featured strongly within the New Curriculum Statement (NCS) – a curriculum that was developed for the new South African education system after 1994, the Revised National Curriculum Statement (RNCS) – a newly developed curriculum meant to streamline and strengthen the NCS in 2001 (Murray, 2012:12), as well as the Curriculum and Assessment Policy Statement (CAPS) – “the third attempt by the Education Department to revise the Curriculum” based on the 2009 Ministerial Task Team’s recommendations (Kanjee & Sayed, 2013:459). It is important at this stage to note some of the reasons why all these reiterations took place. According to Mathonsi (1998) and Nzimande (1995), the old apartheid assessment instruments, such as tests, were used to “deprive blacks from access to resources and opportunities” and to stifle their intellectual development. Some scholars, like Swartz (1992), state that the apartheid system was a deliberate attempt by that regime to apply a system of tests and examinations to regulate blacks’ entry into schools and the economy. Swartz states that this was done through the development of intelligence tests that were used by the State from the beginning to produce theories of intellectual differences between races.

All the different national curriculum reform initiatives that occurred under the auspices of Curriculum 2005 were implemented with the purpose of bringing equity and equality to the curriculum, but certain critics believe that what occurred was far from reaching this objective. For instance, Jansen (2006) refers to these initiatives as mainly

attempts from the new government to purge the apartheid curriculum. From his analysis, one gets the impression that whatever was done by the new government, did not systematically succeed in equalising education in South Africa, but was more about scoring political points. The fact that Curriculum 2005 was hastily implemented did not help the situation either (Jansen, 1999). OBE itself, as well as the process of implementation, were criticised for many different shortcomings. One of the well-documented criticisms towards OBE was its lack of any single historical legacy (Jansen, 2006:322). Jansen states that researchers refer its roots to divergent and unrelated settings. He goes on to say that some connect to “the behavioural psychology associated with B.F. Skinner, others to mastery learning as espoused by Benjamin Bloom, some to the curriculum objectives of Ralph Tyler, yet others claim that it derives from the competency education models associated with vocational education” (Jansen:2006:322) in the United Kingdom. Jansen (2010:323) mentions ten major reasons why OBE was destined to fail:

- the complex language which sometimes was contradicting and confusing;
- the problematic “claims and assumptions about the relationship between curriculum and society”;
- it was based on flawed assumptions about what happened inside the school;
- the anti-democratic assumption of OBE of identifying outcomes in advance for lessons – this violates the epistemology of the structure of some of the subjects in that the creative usage of learners’ knowledge could be contradicted by the already set outcomes, e.g. developing appreciation for a complex writing task in English literature or poetry is something that does not lend itself to instrumental outcomes in advance;
- political and epistemological objections to OBE as curriculum policy; “the focus by OBE on instrumentalism – what a student can demonstrate given a particular set of outcomes – sidestepped the important issue of values in the curriculum” (Jansen, 2010:323);
- “the management of OBE would multiply the administrative burdens of teachers”(Jansen, 2010:323);

- the trivialisation of the curriculum content by OBE;
- teachers who were not well-trained in the assessment approach of OBE; and
- the radical revision that was required for the system of assessment which did not occur.

Other scholars such as Kanjee and Sayed (2013) conceptualise the changes that occurred after apartheid within the Ministry of Education as associations with ministers who took office and key policies on which these phases were based (as briefly stated in Chapter 1) as follows:

Table 2.2: Specific associations with ministers that took office and key policies on which these phases were based

PHASES	MINISTERS WHO TOOK OFFICE	YEARS SPENT IN OFFICE	KEY POLICIES ON WHICH PHASES WERE BASED
Phase 1	Minister Sibusiso Bhengu	1994–1999	The National Curriculum Statement (NCS) and assessment policy in the general and further education and training bands
Phase 2	Minister Kader Asmal	1999–2004	The Revised NCS
Phase 3	Minister Naledi Pandor	2004–2009	The National Policy on Assessment and Qualifications for Schools in the General Education and Training Band
Phase 4	Minister Naledi Pandor	2004–2009	The Foundations for Learning Campaign (FLC)
Phase 5	Minister Angie Motshekga	2009 onward	The National Protocol for Assessment: Grades R–12 and the Curriculum and Policy Statements
Phase 6	Minister Angie Motshekga	2009 onward	The Action Plan to 2014

Concomitant to the above-mentioned changes, a number of shortcomings with regard to educational reform and transformation that were implemented under the watch of each minister can also be noted. Delegates from the National Education and Training

Forum (NETF) that were appointed to deal with issues of reform and transformation in 1992 on the right of the political spectrum wanted to retain certain policies of the past, and delegates on the left wanted changes to be implemented within a short space of time (Jansen, 1999). In the end, according to Jansen (1999), the subject syllabi reviewed by the NETF and its sub-committees reflected the following three main outcomes:

1. Some syllabi such as the African languages remained completely unchanged.
2. Some syllabi such as English and accounting simply added an introductory overview.
3. Some syllabi such as history added new topics and scaled others down (Jansen 1999:63).

Based on the flawed process that was followed in trying to introduce the new curriculum and some of the criticisms that were levelled against the process as well as the new curriculum itself, there seems to be no special attention paid to the proper equalisation of assessments between racial groups, and in particular between examination papers. It is only through the process that was initiated by Umalusi and the DBE that a certain measure to evaluate the quality and the criticism levelled at Grade 12 examination papers was undertaken. This process was undertaken by instituting ministerial task teams. Ministerial task teams that were instituted for this process were commissioned in 1998, 2004 and 2014 successively. With regard to the focus of this study, these task teams established that the language syllabi of the African languages did not promote higher cognitive thinking, they repeated the same material every year, and there was a lack of a common framework of assessment among home languages (DBE, 2014; Ministerial Committee Report (MCR), 1998; Umalusi, 2005). The latest task team report further revealed that certain National Senior Certificate (NSC) HL examination marks were poor predictors of whether or not the learners were likely to pass the NSC overall (DBE, 2014:178).

In order to illustrate the extent of the difference in academic performance between the African Indigenous HLs and that of Afrikaans HL and English HL regarding the end-

of-year examination results, the 2014 ministerial task team issued the statistics displayed in the graph (Figure 2.1).

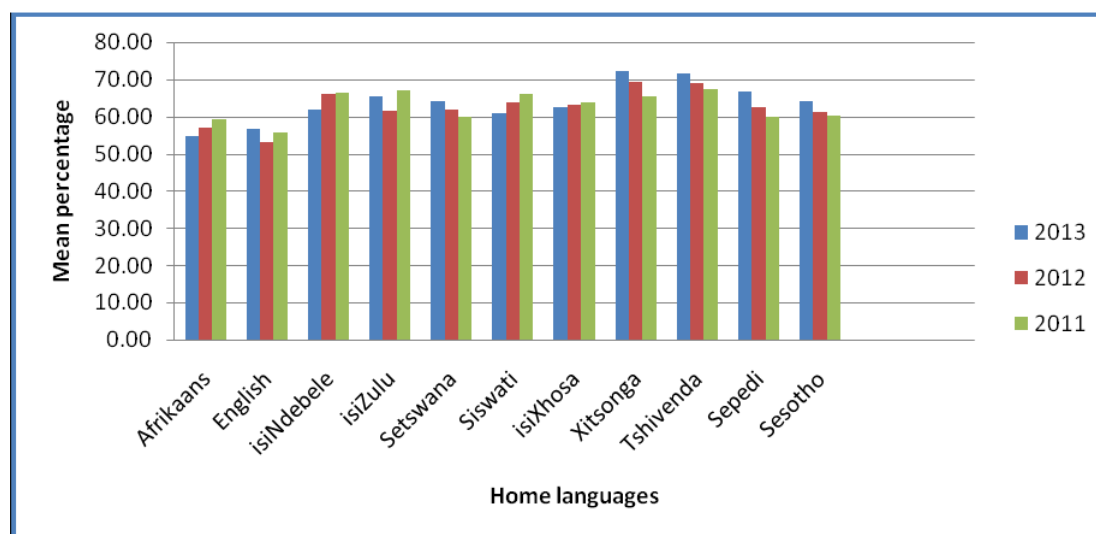


Figure 2.1: Performance of learners by home language 2011–2013 (DBE, 2014:176)

The 2014 task team report (Figure 2.1) indicates that, despite the comparably higher percentage achievement in African languages (although it does not state whether this is statistically significant or not), a high number of learners who wrote the papers in African languages could not pass their NSC examination as opposed to those who wrote either the Afrikaans or English HL papers. For instance, 33,3% of students who wrote the isiZulu, isiXhosa and isiNdebele HL papers who obtained above 60% in these papers, failed the NSC. According to Umalusi (DBE, 2014:178), “the percentage of learners who failed the NSC is represented in terms of the category of pass they achieved in their HL. Just to illustrate this: of the learners who got less than 39% in Afrikaans HL, 80% failed the NSC. Of those who got 60–69%, only 1,3% failed”, of those who got 70–79%, only 0,3% failed and those who got 80–100%, only 0,1% failed. The same pattern is observed for English HL, albeit with a slight difference. This implies that “the chances of failing the NSC if you obtained 60% or higher in Afrikaans and English” are substantially reduced if you compared this with isiZulu, isiXhosa and isiNdebele.

Another area that contributes towards the lack of equality in HL examination papers, that is mentioned by the 2014 ministerial task team, is the lack of a common framework of assessment of these languages. The report indicates that despite an assertion by authorities that a common framework for assessment is being developed, practice in terms of paper format, structure, mark allocation, “level of the texts, and cognitive demand of questions still differ” (DBE, 2014:177). This report concludes by showing that the result of this situation leads to inflated marks for African languages. This could have been avoided as early as 2013 at the standardisation process when advice was given that common standards were necessary for all HL papers, but instead some parties argued that African languages were unique and implied that common standards were not applicable to all HL papers (DBE, 2014:179).

Based on the background to this study and the problems outlined above about the prevailing situation within the South African DBE, it became clear to scholars that a concerted effort was needed to address this situation as is evidenced by Jansen (1998), Kanjee and Sayed (2013), and Lolwana (2005). Umalusi further commissioned a joint undertaking in 2014 between itself and the Inter-institutional Centre for Language Development and Assessment (ICELDA), to create a project to equalise the Grade 12 examination papers of all the 11 official languages of South Africa. In order to address the issue, this study established a research aim in which a model for equalising these imbalances could be addressed. It was decided that a Test of Advanced Language Ability (TALA) be used as a mechanism to close the gap between different languages and bring them to the same level of competency by applying the TALA instrument. The aim was to discard some of the proficiency and oral components of the 300 marks allocated for the final Grade 12 examination paper and replace them with the 60 marks (20%) of TALA. The reason behind the notion of removing the proficiency and oral assessment beyond Grade 10 is the extreme challenge which the department is facing in still ensuring that these components are assessed. However, according to Du Plessis *et al.* (2016), including the TALA in all Grade 12 HL examination papers would create “a degree of overlap and equivalence of a reliable component of assessment which makes it possible to make statistical adjustments

across the total that would further aid comparability and possible mark adjustment in order to ensure fairness". If found to be difficult to implement in the proposed manner, this proposal could be manipulated to accommodate other appropriate ways of incorporating the TALA, even if it means retaining the oral and the proficiency assessment after Grade 10.

The ministerial task team that was instituted by Angie Motshekga in 2014 to investigate Grade 12 examination papers, further indicated that the level of competence that was acceptable for the Grade 12 HL examination was that of the Afrikaans and English papers (DBE, 2014:27). It was, therefore, decided that English be used as the norm for the strategy of equalisation proposed in this study. As indicated in Chapter 1, an important objective of this study is to determine the most appropriate way for the creation of TALAs for the different official languages in South African that are equal across these languages in terms of difficulty as well as what is being tested. Subsequently, it was decided to apply a strategy whereby a translated Sesotho TALA and a newly created Sesotho TALA be compared for adequacy. The aim is therefore to match data accumulated from these tests to determine which one between the two is statistically adequate to be recommended for application in the Grade 12 examination. Following this strategy, it would thus be essential to compare Translated Sesotho TALA and the Newly Created Sesotho TALA the English TALA in terms of adequacy.

2.3 Conclusion

This chapter has shown how the departmental structures of the apartheid era have affected the assessment of subjects in the South African education department. It has done so through a detailed account of how education departments, school systems, grading and assessment bodies were divided. It has also explained how these divisions affected the assessment between different subjects.

It has further explained how, in trying to transform and reform education after 1994, the new South African government came up with a new education system, school systems, grade divisions and assessment bodies. Unfortunately, the new education

system that was introduced by the new government faced a number of challenges, many of which still are not adequately dealt with. Important for this study is that gross inequality still exists in a number of respects between the different home languages in education.

In the next chapter, a literature review about the translation models and methods, as well as strategies and procedures which are appropriate for the translation of the TALA, are provided.

CHAPTER 3: TRANSLATION IN LANGUAGE TESTING

3.1 Introduction

This chapter deals with the relevant literature concerning the translation of the original English TALA into Sesotho. Literature about the most appropriate approach towards the translation of this test are reviewed. This discussion covers the evolution of translation as it relates to testing. It does so by including, among others, a discussion of the most prevalent translation theories and their impact/influence throughout history. It also discusses translation models, strategies and methods that may be employed towards the most adequate translation of the TALA.

The chapter commences by first reviewing information about the development of translation studies into a science and proceeds to discuss the specific models, strategies (also at times referred to as ‘approaches’) and methods that are relevant to this study.

The preceding outline about how this chapter will be structured is based on the notion that translation studies use translation theories (which include models), strategies and methods to explain how translation can be done within a specific context.

The three main concepts mentioned as regards the way in which this chapter will be structured is translation theory (including models), translation strategy and translation method. How and when different translation theories developed will be outlined in the next section titled “the evolution of translation”. These concepts are important in order to inform the choice that one would like to make about the most appropriate approach to be used in the translation of the TALA tests.

3.2 The evolution of translation

Translation theory as a field has been in existence for many years – certain scholars actually date it back to the tower of Babel (Gentzler, 2001:1). However, it was only formalised when it was entered into the “Modern Language Association International Bibliography” in 1983, thereby making it more visible as a field for research in all

aspects of modern language and literature. In addition to this, it is also important to note that it was only during the late 1950s that the debate about the “science” of translation became robust during the time of Noam Chomsky and Eugene Nida (Gentzler, 2001:44). Edwin Gentzler’s *Contemporary translation theories* provides a comprehensive account of the history of translation for both theory and practice. It covers the history of translation theory and translation studies including prominent figures/contributors from North America (encompassing works by I.A. Richards, Ezra Pound, Frederick Will, Lawrence Venuti, Eugene Nida, Noam Chomsky, and others); Germany (which includes research by Wolfram Wiss, Katharina Reiss, Hans Vermeer, and others); the Netherlands and Belgium (incorporating works by James Holmes, André Lefevere, and others); and Russia (including works by Jin Levy, Even-Zohar, etc.). This chapter also focuses on different perspectives on the evolution of translation from different authors so that these divergent views are illustrated. The purpose of this concise historical overview is to situate the current study regarding the appropriateness of the translation theory, strategy and method selected for this particular study.

Two prominent theorists from the early sixties whose work helped to anchor translation as a science are Noam Chomsky and Eugene Nida (Gentzler, 2001:44). Nida studied Chomsky’s theory about his transformational rules and his terminology from his publications, *Syntactic structures* (1957) and *Aspects of the theory of syntax* (1965) and incorporated this theory of syntax into his ‘science’ of translation through his publications, *Message and mission* (1960) and *Toward a Science of Translation* (1964) (Gentzler, 2001:46). Chomsky’s ‘generative transformational grammar’ therefore lent credence to Nida’s ‘science’ of translation.

While the theories of Chomsky and Nida are some of the fundamental theories in the development of translation studies, it should be noted that these theories were not without controversy. There was a debate between Chomsky and Nida about who used grammars for translation first. There is also an ongoing debate between theorists about the validity of Chomsky’s perspective about how the human mind works and unanswered questions about Nida’s claims which were fraught with religious

presuppositions and missionary goals (Gentzler, 2001:46-48). However, these two respective theories made an important contribution to both linguistic and translation theory as we know them today (Bassnet, 2014:6; Gentzler, 2001:47).

Following these theories, another prominent translation theory evolved during the 60s and 70s: Eugene Nida's 'formal' and 'dynamic' equivalence theory (Bassnet, 2014:19). Formal equivalence is also referred to as 'gloss translation' by Nida. Its intention is to assist the reader to understand as much as possible of the source language (SL). Hence, its focus is based more on equivalence pertaining to such issues such as sentence to sentence, concept to concept and genre to genre translation. On the contrary, dynamic equivalence is based on the equivalence effect. This refers to the principle that the effect which the original message had on the receiver of that message should be the same as that experienced by the receiver of the translated message. Another well-known theory belonging to this period is Katharina Reiss and Hans Vermeer's 'Skopos' theory. 'Skopos' in this context refers to "the aim or purpose of a translation" (Vermeer, 2000:221). 'Skopos' theory gave rise to the functionalist approach to translation whose main proponent is Christiane Nord from the late 1980s onwards (Nord, 1997). Following these early translation theories, increasingly more translation theories developed and led to what came to be viewed as contemporary translation theory. On the question about what exactly contemporary translation theory was at the time, the Russian scholar Roman Jakobson proposed that it could be divided into three different fields (Bassnet, 2014:20; Gentzler, 2001:1). According to Gentzler (2001:1) these fields are "intralingual translation, which is a rewording of signs from the same language; an interlingual translation – the interpretation of signs in one language with signs from another language; and intersemiotic translation – the transfer of the signs in one language to another non-verbal sign system". New developments of translation theory after Jakobson produced several 'modern' translation theories such as:

1. The North American Translation Workshop – which is a collection of theories from several translation theorists such as I.A. Richards, Ezra Pound, Frederick Will and Lawrence Venuti (Gentzler, 2001). This category of translators is also

found in other literature such as Bassnet's "Translation: the new critical idiom" (Bassnet, 2014:14). This is a category of translators whom Venuti (1995:17) felt were marginalised by the Anglo-American group simply because they would not like to conform to a notion that a translator should be invisible in his discourse (Bassnet, 2014:14).

2. The science of translation – which is based on linguistics and pioneered by Noam Chomsky and Eugene Nida (including theorists like Wolfram Wilss, and German Functionalists, Katharina Weis and Hans Vermeer).
3. The early translation studies – these were studies which applied the inclusive approach between literary and non-literary camps and proposed a movement towards the formulation of a 'theory of translation' which was frowned upon prior to this.
4. Polysystem theory – a term that was introduced by Itamar Even-Zohar "for the aggregate of literary systems, from high or canonised forms such as poetry, to low or non-canonised forms" (Gentzler, 2001:103). This term was used to define norms that influenced translation decisions.
5. Deconstruction/affirmative productivity – an approach to translation that challenges any fundamental notions that exist in different theories of translation, e.g., "deconstruction resists systems of categorisation that separate source from target text or language from meaning, denies the existence of underlying forms independent of language" (Gentzler, 2001:147).

Besides the 'modern' translation theories outlined above, other authors such as Susan Bassnet, instead of chronologically listing translation movements, claim that translation has always fallen into two categories, namely, the pessimistic and the optimistic (Bassnet, 2014:12). The pessimistic, according to her, are the theorists who see all translation as doomed to failure and emphasise what is lost; while the optimistic acknowledge the difficulties faced by translators, seek solutions and view translation as a fundamental means of enriching a literary system.

Bassnet uses Vladimir Nabokov as an example of a pessimistic theorist to illustrate the nature of how her concept of a pessimistic view works. Nabokov, according to

Bassnet (2014:12), criticised limitations which a translator faces when translating works of art strongly rooted in the source culture of the writer. Nabokov criticises the limitation of a translator to possess a whole range of ‘encyclopaedic’ knowledge about a source culture and being able to combine it with the ability to write competently when executing a literary translation such as his translation of a novel in verse by Alexander Pushkin called ‘Eugene Onegin’ (Bassnet, 2014:12). While he criticises the inability to possess this knowledge, he still recommends that detailed footnotes should be supplied with every translation. On the other hand, Bassnet explains the optimistic category by using *Die Aufgabe des Übersetzers* (“The task of the translator”) by Walter Benjamin as an example. Benjamin views translation as an act of producing a third space where a third language different from the one contained in the source language (SL) and target language (TL) is present. He says that a third language which is pure and unifying is distinct from both the SL and the TL and is revealed by translation (Bassnet, 2014:12).

In contrast to the evolution of translation as portrayed by Gentzler and Bassnet above, Anthony Pym has a different perspective on the history of translation (Pym, 2014:1-4). Pym views this history as emanating from James S. Holmes’ ‘perspective on many things’ in which Holmes proposes ‘things’ that can be done in translation. The so-called ‘things’ here imply categories that had long been used to judge translation. All these things that Holmes conceptualised are featured in Holmes’ conceptual scheme of translation of 1972. He divided his scheme into categories of what were believed to be important in translation before his time and that which he thought were significant in his time (thereby designing what can be called a historical map). Pym (2014) believes that Holmes’ scheme was simple, scientific, and hierarchical. However, Holmes’ scheme or map evolved over time. Figure 3.1 depicts Holmes’ scheme as it appears in Pym (2014:2).

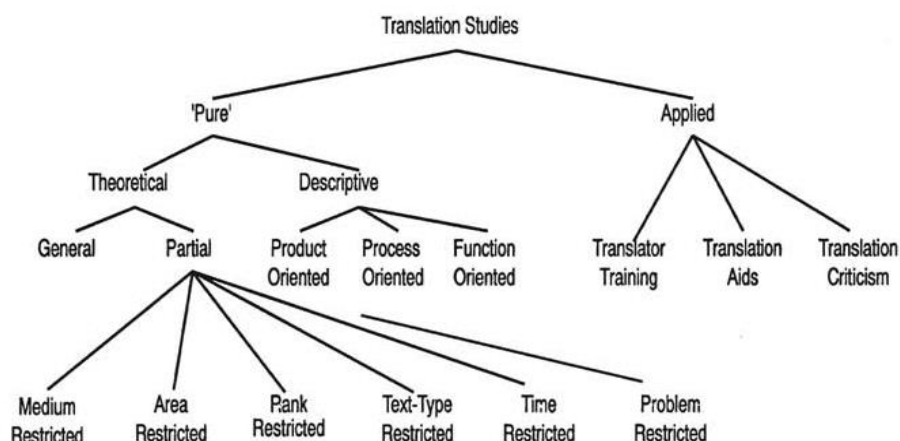


Figure 3.1: “Holmes’ scheme of translation concepts from Pym” (2014:2)

In Holmes’ scheme depicted in Figure 3.1, a concept such as ‘applied’ was opposed to ‘pure’. ‘Pure’ was divided into ‘theoretical’ and ‘descriptive’ (which implied that translation was about theoretical matters and descriptive matters). ‘Descriptive’ was further divided into ‘product oriented’, ‘process oriented’, and ‘function oriented’, and so on (Pym, 2014:2). With regard to issues of a descriptive nature as portrayed here, one understands that translation was meant to deal with issues whereby a function produced certain results as required by a certain process. All these concepts are categories of things that were believed to be important in translation before Holmes’ time and things that he thought were significant in his time. With the above-illustrated map, Holmes implied that the manner in which translation was carried out in the past could be studied according to his product-oriented branch as described in the map. However, this map did not take the historical functions of translation into account. This scheme received considerable attention from scholars who thought it was an important point of departure for developing the history of translation. Pym (2014:5) asserts that almost every part of the scheme can be related to historical events of translation, but concludes by saying that this map is no longer a reliable guide about the history of translation. Confirmation that Holmes’ map of things that can be done in translation served as an important step forward for translation theory, could be seen in the further

development of this historical branch by theorists such as Lambert (1991), Snell-Hornby (1991) and Toury (1991).

As a result of Pym's assertion above, he provides his own map for describing the history of translation and states that translation history can be subdivided into at least three areas, namely, translation archaeology, historical criticism, and explanation.

According to Pym, the term translation 'archaeology' as used in this context is meant to denote the work done by a researcher to establish "who translated what, how, where, when, for whom and with what effect?" (Pym, 2014:5). One could see that answering these questions could lead to a historical account that can cover a wide range of aspects similar to what Gentzler has done with the five categories of modern translation theories. Additionally, the term 'historical criticism' is meant to denote a criticism of approaches that were applied in the past towards translation as well as a comparison between past theories and contemporary ones. A good example of this perspective is the issue of Bassnet's pessimistic and optimistic translations as explained before. Lastly, the term 'explanations' denote why 'archaeology' and 'historical criticisms' occurred. It is concerned with the causation of such information (Pym, 2014:5).

The translation theories of the sixties, seventies and eighties outlined above developed under very difficult circumstances as there was no harmony between literary theorists and theorists supporting translation studies. At the same time, there were failures to observe certain rules between linguistic and literary investigators as already mentioned. There was little dialogue among the representatives of the literary and translation theory camps which led to scholars working in isolation. However, the nineties, according to Gentzler, experienced a proliferation in theories based on translation studies (Gentzler, 2001:187).

A number of new translation theories emerged in the nineties. Scholars began to close the divide mentioned above and engaged more in dialogue. For instance, Gentzler states that Daniel Weissbort in the United States, editor of *Modern poetry in translation*

added Romy Helen, the author of *Translations, poetics, and the stage : six French hamlets* to his editorial board (Gentzler, 2001:188). Through this dialogue, new theoretical models could be used that were interchangeable between the various camps.

Gentzler (2001:190) observes that theory also returned to prominence after a decade of focusing on the descriptive branch of translation studies and polysystem theory. This initiative was taken by the journal “*Target*” between 1989 and 1990. As a result, there was dialogue taking place between the systems theorists and functionalists, between target-text models and transfer-text models through the initiative of the “*Target*” editors. It is understood that this dialogue has broken down the internal barriers, e.g., the notion that the polysystem branch manipulated the field of translation and that this branch was based on dogmatism in that it was preoccupied with the descriptive method and producing merely elitist literary history, in the field of translation.

Through the acceptance of each other’s contributions, many models and methods emerged from prominent theorists such as André Lefevere and Susan Bassnet. However, what is clear from the above discussion of the evolution of translation is that translation theory is borne from many different perspectives. One of the emerging scholars of the nineties is, according to Gentzler (2001:189), Douglas Robinson. Robinson (1991) felt that literary translators were not afforded the same prominence as translation linguists in terms of the development of translation theory. Robinson’s provocative questions which challenged translation theorists and his continual checking of theory against practice made him popular among the practising translators in organisations such as the American Translators Association (ATA) (Gentzler, 2001:189). Because of his influence and that of other translation linguists and literary linguists, a healthy dialogue emerged between these parties. This dialogue resulted in new insight from developments in text typology and functional linguistics. More recently, other disciplines such as sociolinguistics, pragmatics, psycholinguistics, and discourse theory are further contributing to translation theory. This, in turn, helps with the development of new translation methods and models. Sociolinguists, for example,

38

are exploring issues of how register and dialect affect translation and in terms of the relationship of language to social roles and the impact of power and status on communication status (Gentzler, 2001:189). Gentzler goes on to explain how pragmatists, on the other hand, are exploring the effect of speech with relation to information that is being communicated and the performance of this information (i.e., questions about whether different cultures have the same speech acts).

The discussion above provided a historical account of the development of translation studies as a field. However, since history is not a static phenomenon, new dimensions in translation kept on developing, especially during the past twenty years. While the above account might have outlined a compact evolution of translation studies, issues of culture and ideology still have a strong influence on translation studies. With regard to culture, Munday (2012:193) sees translation studies as having been influenced by cultural factors such as the re-writing (the term that was coined by Lefevere [1992]) of literature, translation and gender, and translation and post-colonialism. Translation as rewriting – according to Munday (2012:193) – implies that authorities in charge of translation work, historiography, anthologisation, criticism, and editing are reinventing the wheel in the manner in which they discharge their roles. These people, according to him, exercise their power to influence “the reception, acceptance, or rejection of literary texts” (Munday, 2012:193). He also sees this rewriting as embedded in the ideological stance of authorities in charge of translation work, historiography, anthology, criticism, and editing.

Another cultural issue that emerged over the past twenty years focuses on translation and gender. Related to this, certain feminist theorists see translation work by women as being treated as inferior to original work and feel that work by women is repressed by the public (Munday, 2012:198). Proponents of this view are among others, researchers such as Gordad (1990), Simon (1996) and Von Flotow (2016).

The last prominent cultural issue focuses on post-colonialism. This issue normally covers studies of the history of the former colonies, studies of European powers that were powerful in their time, resistance to the colonist powers, and imbalances of power

between the colonised and the coloniser. Authors such as Niranjana (1992) and Bassnett and Trivedi (2012), feature prominently with regard to issues such as the power imbalance between languages that translation studies might have not considered, issues of Western translation theory that might be flawed with regard to notions of texts and meaning, and the fact that translation in the colonial context builds on the image of domination by Western philosophy (Munday, 2012:204).

New technology has also had an impact on the evolution of translation studies. Media engenders issues of machine translation as well as audio-visual translation. All these issues are impacting future developments of translation studies. For instance, machine translation has led to a rapid development of translation studies (Bassnet 2002:16). Moreover, audio-visual translation has helped to improve subtitling approaches.

The following section presents relevant theories, strategies and methods from which concepts pertinent to this study may be derived. These issues – some of which are found in the historical account above – are relevant for this study since they illustrate how different techniques of translation could be applied in the study.

3.3 Nord's functionalist approach to translation

With regard to the selection of a particular translation model as explained in the introduction to this study, several studies have investigated test translation within the framework of Nord's model of translation (Nord,1997), which follows a functionalist approach, which is just a strategy for the purpose of this study. This model has yielded desirable outcomes as regards the translation of language tests similar in design to the TALA. It was used by Butler (2017) for the study, "Translating the Test of Academic Levels into Sesotho" and by Van Dyk *et al.* (2011) for the study, "Levelling the playing field: an investigation into the translation of academic literacy tests". The following section will provide a justification for why this model is suitable for this study. The practical application of test translation will be provided in Chapter 5 which deals with the methodology of this research, specifically in the section, "Preparation of stimulus material".

Regarding the abovementioned studies, it is argued that desirable outcomes were achieved due to the focus on the functions and the act of translating the text. In the functionalist approach, text analysis is focused on answering the question about the function of the text, not just any function but the dominant function as espoused by Katharina Reis (Nord, 1997:39). Determining the dominant function of the text before embarking upon the translation of the text and making sure that this function is achieved in the translated text, is prominent in the current study. This notion ties with Vermeer's 1989 Skopos theory which deems the purpose (Skopos) of the text as that which is called the dominant function by the Functionalists. However, it should be noted that the Skopos of the source text, according to Vermeer, is less prominent in Skopos theory than in equivalence-based theories (Nord,1997:12). It is important to note this point as no assumption should be made that the Skopos theory will cover issues of equivalence in this study.

The functionalist approach originated from the notion that translation performs a combination of communicative functions. Nord mentions four of these functions, three of which she accepts originate from Karl Buhler's organon model (Nord, 1997:40). This model includes the following three functions: referential (where the referent might be known or unknown to the translator, such as being a traffic accident, "language or specific use of language", a 'correct way of handling a machine' (Nord. 1997:40), etc.); the expressive (where the referent is either sender-oriented or receiver-oriented [attitude or opinion] such as when an Indian man compares his wife's eyes to that of a cow and it is appreciated in India but scorned in Germany); and the appellative (where the referent is an appeal to our emotions about how we express ourselves). The fourth function is adapted from Roman Jakobson by Nord (1997) and is the phatic [where the referent is a linguistic, non-linguistic and paralinguistic expression "to establish, maintain or end contact between the sender and the receiver"]

(Nord,1997:40)]. What follows next, is a ¹discussion about how each of the above-mentioned functions is applied.

Referential function: In the instance where the referent is a language or specific function of a language, the text function may be metalinguistic. For instance, if a question in the TALA requires a candidate to identify a linking word denoting cause and effect, a word such as *hobane* (because) and a phrase like *ka lebaka leo* (therefore) could be used in a Sesotho translation of such items.

Expressive function: As regards this function, one could refer “to the sender’s attitude toward the objects and phenomena of the world” (Nord, 1997:41). As way of illustration, one may compare the example used by Nord (1997:41) with what the Sesotho version would look like. Nord (1997:42) uses Simone de Beauvoir’s *Une mort très douce*, the adjective *douce* (sweet) as expressing an emotional experience of a dying person. “The English translation, *A very easy death*, expresses a kind of evaluation, perhaps as seen from a doctor’s point of view” (Nord, 1997:42). The German translation, according to Nord (1997:42), “*Ein sanfter Tod*, combines the two aspects because *sanft* might mean ‘sweet’ from the dying person’s point of view and ‘easy’ or ‘painless’ from a more detached perspective”. Comparing this example to a possible Sesotho version: the same translation, *lefu le thotseng* (death without any drama), from both a dying person’s point of view and a detached point of view is possible.

Appellative function: This function relates to inducing a response in the receiver by using linguistic or stylistic devices. According to Nord (1997:43), “the source text normally appeals to source-culture susceptibility and experience while the appellative

¹ In instances where an immediate example from the TALA is available, an example from the TALA will be used, but in instances where such an example is not readily available, an example from primary sources and/or other sources will be used. It must be noted that not every example will be available from the TALA.

function of a translation is bound to have a different target". If, for example, the term 'minority group' in America would affect black people in a certain way, in South Africa it could affect white, coloured, indian, or the Khoi-san in a similar way. Or, if a source culture in America appeals for clean energy from water, then in Botswana it would be for either the sun or the wind (because Botswana is a semi desert).

Phatic function: Nord (1997:44) states that this "function relies on the conventionality of the linguistic, non-linguistic and paralinguistic means such as small talk about the weather or a conventional proverb used" – which might be unconventional in another language or 'peg' in tourist information text just to friendly humour a visitor. For example, translating a piece of greeting text from Sesotho or any other indigenous African language to any of the European languages could be laborious for a translator of European origin. The translator may, therefore, just translate it briefly as "they greeted", or something close to that. However, from an African point of view it could be lacking valuable information about what people actually said to each other when they greeted, irrespective of its laboriousness. The reason for this is that Africans spend a lot of time greeting each other and even go to the extent of asking each other about life conditions at home as well as weather conditions. For example, a simple greeting like, "How do you do" and responded to as "How do you do" in English, could be articulated as, (*Le kae?*) / How are you /, in Sesotho and be responded to as, (*Re a phela ha re tsebe lona. Ha e se letsatsi lena le baletseng. Ha re tsebe hore na re tla kotula selemong sena*) / We are still alive, we do not know about you. Except this drought that we are experiencing. We are not even sure if we will harvest anything this year.

The functionalist approach as explained above will feature strongly in the current study, for the specific reason that the overall function of the translation in this study is that of translating an assessment instrument such as the TALA. In addition, the specific evaluative functions of the different subsections of the original English TALA as well as the Sesotho TALA will have to be kept in mind in the translation of the texts and questions based on such texts. The functional approach will further be helpful in

the analysis of linguistic and cultural differences between the two tests and their subtests.

In addition to translation approaches (such as Nord's functionalist approach illustrated above), translation models and methods help to produce strategies for translation studies in general. The functionalist approach to translation has partly introduced the concept of a translation model where it referred to Karl Bhuler's organon model above. How that model operates and can be applied to some of the aspects of the translation of the TALA was demonstrated in the discussion above. Some strategies are more inclined towards certain theories, models and methods than others. A review of prominent theories, models and methods relevant to this study will be done in the remainder of the chapter in order to explain how these strategies could be utilised in this study. In order to explain how the mental process of translation work, it is important to examine Gile's sequential model of translation (1995) before other models, which are more application based, are examined. This model serves as an introduction to translation theory and it explains how a translator thinks about a translation and should not be confused with other models which are applied during the actual translation process.

3.4 Gile's sequential model of translation

This model is based on the sequence of steps which the translator follows when engaging in translation activity. According to this model, a translator's focus when translating is based on two phases, viz, the comprehension phase and the reformulation phase.

The comprehension phase: This is the phase in which the meaning hypothesis is conceptualised in translation. This is done through an attempt by the translator to discover what the meaning of the translation unit is. The translator discovers this meaning by using his pre-knowledge called the knowledge base in this model as well as the knowledge gained from sources referred to as knowledge acquisition. The translator, according to Gile (1994:103), runs a series of plausibility tests to verify the

translation unit until he reaches the best plausible unit. This phase of Gile's model will form part of conceptualising the content of the translator training workshop for translators that will be used for this study.

The reformulation phase: Reformulation of the translation unit takes place when this unit is verbalised into the TL. In order for this to happen effectively, a series of plausibility tests is once more performed with the TL to ensure translation fidelity. This is done by ensuring that none of the SL information is omitted in the translation. The translator checks this by checking language correctness, stylistic appropriateness and terminological acceptability. This is done continuously until the results of the translation for that unit are acceptable to the translator. This process ensures that the quality of translation is high and no error is left unattended. It can also be done during the analysis of the translated text to ensure the quality of the translation. This approach as well as back-translation is beneficial for the analysis of errors in the translation of the TALA with which this study is concerned.

First among the application-based models of translation to be discussed is Vinay and Darbelnet's model (1995). Vinay and Darbelnet's model makes a considerable contribution in terms of procedures and strategies that can be applied to the translation of the English TALA.

3.5 Vinay and Darbelnet's model

This model is based on the analysis of translated English and French texts. Vinay and Darbelnet (1995) focused on the difference between the strategies and procedures which are applied when a translation practitioner engages in translation. They illustrated this difference through a comparative analysis between English and French texts. Their model gives a clear definition of the method of translation as well as the difference between strategies and procedures of translation.

In their model, the first step in the method of translation is that a translator must establish the relationships between the specific manifestations of two linguistic systems. These linguistic systems, according to them, are the linguistic system that

was used in the ST and the one that was used in the TT which is still potential and adaptable (Vinay & Darbelnet, 1995:30). This engagement, according to them, gives a translator a fixed starting point in his mind. This helps a translator to decide what target he wants to reach based on his reading of the ST. The SL reflection must help the translator to focus on the message of the TL. By doing so, the translator should be able to identify or discover whether SL units or sentences have been omitted. This method uses two general strategies and seven procedures.

According to Vinay and Darbelnet (1995), strategies refer to the overall orientation of the translator and procedures are the specific methods used in a translation. Two primary strategies, a direct and an oblique strategy, feature strongly in their model. These two strategies could be equated to the notions of 'literal' and 'free' translation. Literal translation is referred to as a synonym for direct translation (Munday, 2012:86). Direct translation, according to this model, refers to a strategy where message elements of the SL can be directly transposed into the TL without changing the message. This happens because this transposal is based on either structural parallelism or on metalinguistic parallelism. Structural parallelism refers to instances in translation where the syntactic order is carried from ST to TT without causing any stylistic difficulty. Metalinguistic parallelism relates to occurrences where units of speech of the ST are transposed to the TT. Gaps or lacunae can still exist in the SL even where structural parallelism or metalinguistic parallelism have occurred, and when this happens translators may need to take certain precautionary measures of stylistic nature to address the situation so that the message is the same for the two languages.

It may happen as well that sometimes, based on metalinguistic and structural differences, "stylistic effects cannot be transposed into the TL without upsetting the syntactic order" of elements of the message or the lexis (Darbelnet, 1995:32). When this happens, complex and unusual methods need to be applied. These procedures are called oblique translation methods.

The direct translation strategy entails three possibilities, according to Vinay and Darbelnet (1995:32). These options are borrowing, calque and literal translation. In order to overcome a gap, especially a metalinguistic one, borrowing is the simplest procedure in that the SL word is directly transferred to the TL.

3.6 Borrowing

Borrowing, according to Vinay and Darbelnet (1995:32), can be used to introduce a ‘flavour’ of the SL into the TL. However, for translation from English into Sesotho the purpose for using borrowing could have an added objective. It has been a tradition in Sesotho to use borrowing where the Sesotho language lacks an equivalent word or concept. Even though this could be the primary objective, it will still introduce a flavour of the SL into the TL. There is extensive use of borrowing in the translation of the original English TALA into Sesotho, both for part 1 and part 2 of the test. Each part of the TALA is identified by the title of the text upon which it is based. Part 1 of the English version of TALA is based on the text titled “Foods that can save the world”, and part 2 of the English version is called “King of the carbs”. As already stated, the lack of equivalent words or concepts in Sesotho is one major reason why borrowing is utilised in this study as indicated in Tables 3.1 and 3.2. Only words that have been borrowed will be shown in the following two tables.

Table 3.1: Examples of borrowing from TALA part 1

ENGLISH	SESOTHO
/rice/	(rice)
/diabetes/	(diabetes)
/scuba/	(scuba)

Table 3.2: Examples of borrowing from TALA part 2

ENGLISH	SESOTHO
---------	---------

/ratio/	(ratio)
/monosodium glutamate /	(monosodium glutamate)
/ index /	(index)

3.7 Calque

A **calque**, according to Vinay and Darbelnet (1995), is a “special kind of borrowing where the SL expression or structure is transferred in a literal translation”. All the elements of a borrowed expression are translated literally in the TL. The results of this are a lexical and a structural calque. A lexical calque is defined as a calque that respects the syntactic structure of the TL and a structural one introduces a new construction into the TL. Each of these calques will be illustrated in separate tables below. Among the different types of calques there may be the ones, like borrowing, that have become an integral part of the TL language after a long time. This makes them fixed calques. There are many instances of fixed calque between English and Sesotho (examples from the TALA are provided in the tables below).

Table 3.3: Examples of calque from TALA part 1

ENGLISH	SESOTHO	TYPE OF CALQUE
/Europe/	(<i>Yuropa</i>)	Structural (Sesotho words never have a ‘y’ in their structure, y that we see is evolving from ‘ea’)
/hectares/	(<i>dihekthare</i>)	Lexical & fixed
/tonnes/	(ditone)	Lexical & fixed

Table 3.4: Examples of calque from TALA part 2

ENGLISH	SESOTHO	TYPE OF CALQUE
/column/	(<i>kholomo</i>)	Lexical & fixed

/graph /	(kerafo)	Lexical & fixed
/cramps/	(<i>dikerempe</i>)	lexical

3.8 Literal translation

According to Vinay and Darbelnet (1995:33), **literal translation** is a well-known method of translation. They emphasise that it is a method where the direct transfer of an SL text of both grammar and idiom is effected in the TL appropriately. They further outline the method of doing this correctly. For instance, they indicate that it is the most unique reversible solution and complete in itself. By this, they imply that a literally translated text can be reversed to the ST unproblematically. As Vinay and Darbelnet (1995:33) state, this is most possible between languages of the same family and culture such as French and English because of their common metalinguistic concepts and coexistence. They also indicate that it is quite common between languages which belong to the same family such as English and French. This can, however, not be said for English and Sesotho which are the two languages with which this study is concerned. Sesotho and English do not belong to the same language family in terms of culture and metalinguistic concepts. The Sesotho culture is rooted in ancient African culture, whereas the English culture that is used is modern and progressive. Metalinguistically, Sesotho is disjunctive whereas English is conjunctive. For example, note how disjunctive the underlined Sesotho verb is from the underlined conjunctive English verb in the following sentence:

/English: Man walks./

The verb in the above English sentence is conjunctive because it is made up of two bound morphemes.

(Sesotho: Monna o a tsamaya.)

The verb in the above Sesotho sentence is disjunctive because is made up of three separate morphemes.

For an oblique strategy there are nine procedures that can be used in order for a translation to be successful (Vinay & Darbelnet, 1995:33). These procedures are: transposition, modulation, equivalence, adaptation, amplification, false friends, loss, gain and compensation, explication, and generalisation.

3.8.1 Transposition

According to Vinay and Darbelnet (1995), transposition involves replacing one word class with another without changing the meaning of the message. This is a process that Munday (2012) explains as exchanging one part of speech for another (e.g. noun for verb) without changing the meaning. Transposition can either be obligatory or optional. This often occurs between English and Sesotho between nouns and infinitives. An infinitive in English is often translated into a noun belonging to noun class 15 in Sesotho. Based on a sentence from TALA part 1 in English, the following translation is observable:

English: /scientists have worked **to create** foods that can boost health/

Sesotho: (*boramahlale ba sebeditse **ho etsa** dijo tse ka ntlafatsang bophelo*)

Back translation: /scientists have worked to make foods that can improve life/

An infinitive **/to create/** in the English version of the sentence is translated as a noun (**ho etsa**) from noun class 15 in Sesotho.

3.8.2 Modulation

This procedure, according to Vinay and Darbelnet (1995:36), is applied through the variation of the form of the message by changing the perspective or the semantics of the SL. This change, according to Vinay and Darbelnet (1995), is justifiable only if the translation results in a grammatically correct utterance. This justifiability can happen even if a literal or a transposed translation had occurred. Modulation can also either be obligatory or optional. Munday (2012:88) provides clear examples of instances where this can occur. An example of obligatory modulation is as follows in French:

“the time *when* translated as *le momento ou* (lit. ‘the moment where’)”

An example of optional modulation is the following: “though linked to preferred structures of the following languages: e.g. the reversal of point of view in it is *not difficult* to show”

1. *Il est facile de demontre* (lit. ‘it is easy to show’) (Munday, 2012:88).

From the TALA, the following occurrence of modulation can be observed:

English: /Blame it on the potato/

Sesotho: (*Tshwaya tapole phoso bakeng sa sena*) (Blame the potato for this)

It is important to note that the above type of construction occurs frequently between English and Sesotho.

3.8.3 Equivalence

According to Vinay and Darbelnet (1995:38), this concept refers to the use of different stylistic and structural methods where the aim is to produce equivalent texts. These texts are equivalent in that even though they do not say the same thing, they are not different with regard to describing a process, situation or the object. These authors also claim that this type of translation happens where a so-called bilingual population exists. A good example to illustrate this situation is the case of Isizulu and Sesotho, which are two South African languages spoken by many people in the Gauteng Province and the Free State Province in South Africa. Compare, for example, the following proverb in Isizulu and its Sesotho translation:

Isizulu: /*Likhipha umkhovu etsheni*/ (The sun exhumes a mythical manlike creature from under a tree).

Sesotho: /*Le ntsha kwena bodibeng*/ (The sun removes a crocodile from under the water).

English equivalent: (As hot as fried hell).

It is clear from the translation of this proverb from Isizulu into Sesotho that these translations do not say the same thing, but the message conveyed is the same. A classic example given by Vinay and Darbelnet (1995) is a situation where an amateur carpenter accidentally strikes his finger with a hammer. If he were French, he would cry “Aie!”, but if he was English he would cry “Ouch!”. These authors indicate that this method of creating equivalence is often applied in the translation of idioms and proverbs.

From the TALA, the following example can be used to illustrate this:

English: /this imported foodstuff **became a peasant favourite**/

Sesotho: (*sejo sena se neng se tswa dinaheng tse ka ntle, e ile ya e ba thatohatsi ya ba kollang ntsi hanong.*) (lit. it became a favourite for those who struggle to catch a fly in order to put it in their mouths)

3.8.4 Adaptation

Adaption is referred to as the extreme limit of translation by Vinay and Darbelnet (1995:39). The procedure of adaptation is often considered appropriate for the translation of language tests because adaptation helps with the objective transference of the original construct of the test into the TL. This section will define adaptation in the context of Vinay and Darbelnet’s model first and in the context of Koch’s model later on in the chapter. Vinay and Darbelnet define adaptation as the translation procedure where a cultural reference is changed when a situation in the source culture does not exist in the target culture. In situations like these, the translator has to create a new situation that can be considered equivalent to that of the SL. These authors refer to adaptation as a ‘special kind of equivalence’ or ‘situational equivalence’. A simple example could be that of an English father who would never think of kissing his son on the mouth, but something which a French father would regard as normal behaviour. If a sentence occurs in French saying: “A father kissed his **son** on the mouth”, it is possible that it could be translated as: “A father kissed his **daughter** on

the mouth” in English. The onus in this extreme illustration lies with the translator if he would like to transfer the cultural meaning or not.

Adaptation is used extensively in the current study. For instance, in order to create a culturally equivalent context based on the English language, which does not exist in Sesotho, translators resorted to using a lot of borrowing from English. The rationale behind the notion is that this is a form of adaptation based on Darbelnet’s idea that borrowing creates a flavour of the SL in the TL (Vinay & Darbelnet, 1995:32). Take the heading of the main text, “King of the carbs” of the English TALA part 2, for instance. This text is strongly embedded in the culture of the English language, both lexically and figuratively. The heading was translated as /*Morena wa dicarbohydrate*/ (lit. King of carbohydrates) in the Sesotho TALA part 2. Lexically speaking, it would have been far-fetched to translate carbs as /*dicarb*/ as the word carb is an informal word for carbohydrates which some of the English speakers might not even be familiar with, let alone the Sesotho speakers. For this reason, a version closer to the English word that would be acceptable and understandable to Sesotho speakers was used. However, figuratively speaking, the whole heading as metaphor is culturally far-removed from the Sesotho language. Carbohydrates (food in Sesotho is not reduced to this finer component) is neither a cultural nor a regular topic in Sesotho. However, to close the gap between the two languages and questions based on this text for learners, it became necessary to be translated as such. This rationale was applied extensively through the course of both parts of the Translated Sesotho TALA in this study.

3.8.5 Amplification

Amplification refers to a procedure where the TL uses more words, often because of syntactic expansion. Because a language such as Sesotho is disjunctive and lacks many of the modern terminology, it uses amplification extensively. A disjunctive language such as Sesotho makes use of more additional morphemes for different parts of speech than languages such as English. Take, for example, a typical sentence from the English TALA part 1 and the Translated Sesotho TALA part 1:

English: (food baddie or misunderstood enhancer?)

Sesotho: /*mphato wa dijo kapa motlatselleltsi ya sa utlwisisweng*/

Comparative analysis of the above sentence segments

(food baddie)>/mphato wa dijo/

(or)> /*kapa*/

(misunderstood enhancer)> /*motlatselleltsi ya sa utlwisisweng*/ > (supplemental helper)

A lack of terminology also often leads to certain terminology from English being paraphrased, thereby using more Sesotho words to achieve better clarity. Note the following illustration from the English TALA part 1 and the Translated Sesotho TALA part 1:

Eng>SeSo: (Gene)>/*phatsa ya lefutso?*/ > (inherited particle)

Eng>SeSo: (super)>/*matla a fetang a tlhahol*/ > (beyond ordinary power)

3.8.6 False friends

False friends involve a procedure of identifying structurally similar terms in the SL and TL that deceive the user to think that the meaning is the same, e.g. the French *librarie* is different from English 'library' in that it means bookstore. Also, in Sesotho, the word *buka* is different from the English 'booker' in that it means book.

3.8.7 Loss, gain and compensation

This refers to a situation where certain words get lost or gained during the translation of the ST, e.g., if a (magician)>/*maselamose*/ becomes a villain during the course of a text a translator is compelled to switch 'magician' to an evil magician /*moloil*/ in Sesotho.

3.8.8 Explication

Implicit information in the ST is rendered explicit in the TT. The following clause from the English TALA part 1 and Sesotho TALA part 1 can serve as an example of explication:

English: (when it is covered by water for too long it becomes **dormant**)

Sesotho: *Iha e pupeditswe ke metsi ka nako e telele haholo e a itjhwesa/*

Back translation: (if it is under the water for longer, it fakes death)

The way in which the word 'dormant' is translated is more explicit in Sesotho than English as it implies that the plant pretended to be dead for some time.

3.8.9 Generalisation

This refers to the use of a more general word in the TT instead of a more specific word. For example, the title of the English TALA part 1 (For the love of strawberries) could be generalised as */Bakeng sa lerato la menokotshwai/* (for the love of berries).

The example above would have been a more general translation than when we say, */Bakeng sa lerato la di-strawberry/* since *menokotshwai* is simply a general term for berries. The latter translation used borrowed terminology for the strawberry in order to be more specific than the former. The former uses a general but more cultural term for strawberry than the latter which is borrowed, but more specific.

The above model of translation by Vinay and Darbelnet is enlightening with regard to helping the researcher to make an informed choice with regard to the selection of methods that can be utilised for translation in this study. It also empowers the translator to engage in text analysis as one of the actions necessary to understand the act of evaluating the effectiveness of a translated text.

3.9 Strategies and procedures in test translation specifically

Relevant information on test translation in general, as well as specific information about studies which were undertaken about test adaptation, are explored in the following section. This is pertinent because the present study deals with the translation of tests specifically.

3.9.1 Test translation

This section focuses on translation options which are available for test translation specifically. With regard to developing tests that are available in different cultures and languages, Koch (2009:304) mentions three different options, namely: direct translation, adaptation, and simultaneous development. Koch (2009), however, contends that although direct translation is regarded as the best option and that it is easy and less cumbersome to use, it may not be the most appropriate option for language tests since it produces a lot of discrepancies between the original and the translated texts. Koch concludes that the best option for the translation of language tests is to adapt the test by translating a set of items literally and to change the words or content of problematic items in order to enhance their appropriateness for a new culture and context.

Furthermore, it is important to take note of the procedures that are used in the translation of achievement tests in order to select appropriate procedures for this study. Worth considering are the achievement tests that were translated and adapted for the Third International Mathematics and Science study (TIMMS) by the International Association for the Evaluation of Educational Achievement (IEA) in 1995 (Gierl, 2000). In 1995, the TIMMS test was administered in 42 different languages. The adaptation of renowned international tests such as the TIMMS into 42 different languages indicates the value of test adaptation. The International Test Commission (ITC) guidelines (ITC, 2017:5) also point out that the translation and test adaptation advances have been necessitated by the growing “cross-cultural psychology, large scale international comparative studies” in education achievement, credentialing

which is used in the whole world, and fairness required in testing considerations that includes learners often having a choice of language(s) in which they prefer to be assessed.

3.9.2 Towards using adaptation as a method of translation

In accordance with the notion by Koch (2009), the translation of the English TALA would further be based on guidelines by the International Test Commission (ITC). Koch's (2009) study, "The case for bilingual language tests: a study of test adaptation and analysis", for instance, focuses specifically on the process of adaptation in test translation. Apart from a description of how the process of adaptation could be applied, this study focuses on applying the guidelines of the ITC to the adaptation of the Test of Cognitive Academic Language Proficiency from English into isiXhosa. These guidelines would also be relevant for the present study and will include guidelines D.5., D.6., D.7. and D.9 which were used in Koch's study, but will not be the only guidelines from the ITC which will be considered. In order to illustrate the value of these guidelines, consider the content of the four above mentioned guidelines in Koch's study: D.5. compels test developers to provide systematic judgemental evidence for any adaptation decision which they take; D.6. "requires of them to ensure that the data collection design permits the use of appropriate statistical techniques to establish item equivalence between the different language versions of the instrument; D.7. compels them to apply appropriate statistical techniques" (Koch, 2009:304); and D.9. requires the provision of statistical evidence of the equivalence of questions for the intended populations. These guidelines will be considered for the present study.

A study by Hambleton and Patsula (1998) who conducted research about "adapting tests for use in multiple languages and cultures" is of particular interest with regard to adapting tests into different languages and cultures. In this study, the following principles were outlined for test adaptation:

- (1) consider several advantages and disadvantages of adapting tests from one language and culture to another, (2) review several sources of error or invalidity

associated with adapting tests and to suggest ways to reduce those errors, and (3) consider test adaptation advances in one rapidly emerging area of social research – quality of life measures (Hambleton & Patsula, 1998:156).

Although the principles mentioned above were considered in the context of psychological tests and not in the context of language studies, it is interesting to note that this study outlines clear advantages and disadvantages of test adaptation. For instance, the first advantage mentioned in the study is that “using adapted tests is usually cheaper and faster than preparing a new test and may permit the use of existing test score norms” (Hambleton & Patsula, 1998:155) for the analysis and comparison of the tests. The second advantage is that “national, cross-language, and/or cross-ethnic comparative studies require adapted tests” in order to ensure their fairness (Hambleton & Patsula, 1998:155). The third advantage is that “researchers often have a sense of security which comes from adapting a test rather than initiating a new test development project in a second language and/or culture” (Hambleton & Patsula, 1998:155). The last advantage is that “tests are sometimes adapted to enhance fairness by enabling persons to take tests in their preferred language” (Hambleton & Patsula, 1998:155).

The above study by Hambleton and Patsula (1998) also mentions the following as disadvantages of test adaptation: “first, test adaptation is sometimes done because of the hope or mistaken belief that a valid test will result”, not taking into account that “a researcher using an adapted test still has the responsibility of producing evidence of validity in the context where that adapted test is used”; and second, “researchers risk imposing conclusions based on concepts which exist in their own cultures which are foreign, or at least partially incorrect, when used in another culture” (Ronald *et al*, 1998:156). These observations by Hambleton and Patsula (1998) are important for the translator since this awareness could help the translator not to use adaptation for its own sake only, but for informed reasons.

The aim of the current study is to determine which one between Translated/adapted the TALA and the Newly Created TALA is most adequate for the equalisation of HL

papers. Since test adaptation forms part of the translation strategy it is used in this study and the series of guidelines of the International Test Commission (ITC) for the development or adaptation of tests into more than one language is used as is illustrated in Chapter 5. These guidelines help to ensure linguistic equivalence between different versions of tests. This study also subscribes to the necessity for validating equivalence between tests as well as finding sources of error in different tests as reflected by the 22 ITC points.

There are also other important principles of test adaptation that are utilised in this study. For instance, literal translation is considered part of the adaptation method of translation (Koch, 2009:304). For this reason, a literal translation strategy is followed in this study. According to Newmark (1998:76), a literal translation is the first step of translation. In order to emphasise the importance of literal translation as an integral part of every translation procedure, he goes on to say that a good translator only abandons a literal strategy when it is plainly inexact or badly written (this refers to the TT after due revision by the translator). It is during such instances where other translation strategies such as transposition (changing one part of speech with another), modulation (a change in the semantics or point of view), and equivalence (used, for example, in the translation of idioms) could be implemented (Vinay & Darbelnet, 1995).

Although test adaptation is adopted as the default translation method in this study, it does not exclude other important translation procedures. While widely used with direct translation, a strategy of back-translation is also utilised in this study for the purpose of comparison and correction. Back translation is a crucial strategy in that it enables the translator to make corrections where the source text does not correspond with the target text. The translator can, therefore, be called upon to justify aspects of a translated text if necessary (Newmark, 1998:282). This is done through the translation of the target text (TT) back to the source text (ST) by a different translator in order to compare the original version with the back-translated one. If the translated text does not yield an equivalent version, then adjustments can be effected (Van Dyk *et al.*, 2011). The advantage of this method, according to Van Dyk *et al.*, (2011:157), is that

researchers have a measure of control over the translated end product, especially if these researchers are not familiar with the target language.

3.10 Conclusion

This chapter explored the evolution of translation to its modern status as the discipline of “Translation Studies”. It outlined some of the early translation theories and showed how they interacted with other disciplines such as linguistics, literary theory, and communication studies to formulate what is considered independent translation studies today. It went on to illustrate how different models and methods helped to create useful strategies and procedures for translation studies. All of this was explained in the context of the purpose of this study which is the identification of the most appropriate approaches, procedures and strategies that can be used for the translation of the English TALA into Sesotho.

CHAPTER 4: VALIDITY AS PROPERTY OF LANGUAGE TESTING

4.1 Introduction

This chapter presents an overview of relevant literature on validity in language testing. In particular, the purpose of the chapter is to situate the current study in the broader debates around testing validity. The reasoning behind this is to ensure that the Sesotho TALA be on the same level as the English TALA that was created by the project team to which this study belongs. Theoretically speaking, the validity of tests is important in both translated and newly created tests in order to ensure that such tests are adequate and appropriate for their purpose. The discussion of literature that focuses on the concept of validity will cover its scope in language testing. Moreover, literature on the common underlying construct for the Curriculum and Assessment Policy Statement (CAPS) for languages will be critically reviewed in order to make visible such an underlying construct. In other words, an analysis of sources from the DBE together with sources from the project team to which this study belongs will be integrated in order to theoretically defend the potential usage of the Sesotho TALA in CAPS.

4.2 The scope of validity in language testing

For the benefit of this study, three perspectives are suggested for validity. These include the orthodox, the postmodernist and the contemporary perspectives. The orthodox perspective is well-known and there is no doubt about its place in the order of things as far as the concept of validity is concerned. However, there is uncertainty about whether there are only one or more perspectives following the orthodox perspective of validity. Some scholars subscribe to the “new orthodoxy”, for example, Weideman (2008:87) and Weir (2005a:13) and others address themselves to the “current orthodoxy”, for example Sebolai (2016:215) and Weideman (2017:7), in addition to the orthodox perspective. Furthermore, there are scholars who refer to just the “current perspective” in addition to the orthodox perspective such as Chapelle (1999:255, 258), Du Plessis (2014:2), Fulcher and Davidson (2007:16) and Sebolai

(2016:14,15). Then there are those who subscribe to the perspectives of orthodox and current only. One possible reason for this is that for these scholars, research was conducted at a time when the prevailing perspectives were just the orthodox and current. However, the current study is being undertaken at a time when various prominent scholars talk about perspectives such as orthodox, new orthodoxy, current orthodoxy, beyond Messick and current, just to name a few. For this reason, the researcher for the current study decided to conduct an analysis of perspectives that exist currently. It was, therefore, decided to divide them into three different perspectives as mentioned at the beginning of this section. The reason for referring to the perspective succeeding the orthodox as “postmodernist”, is based on the notion that this perspective turns around tradition and reconstructs the widely accepted theory of what orthodox validity entails. The third perspective according to this study, namely, the “contemporary perspective”, is self-explanatory as it deals with present information and reasoning.

4.2.1 The orthodox perspective

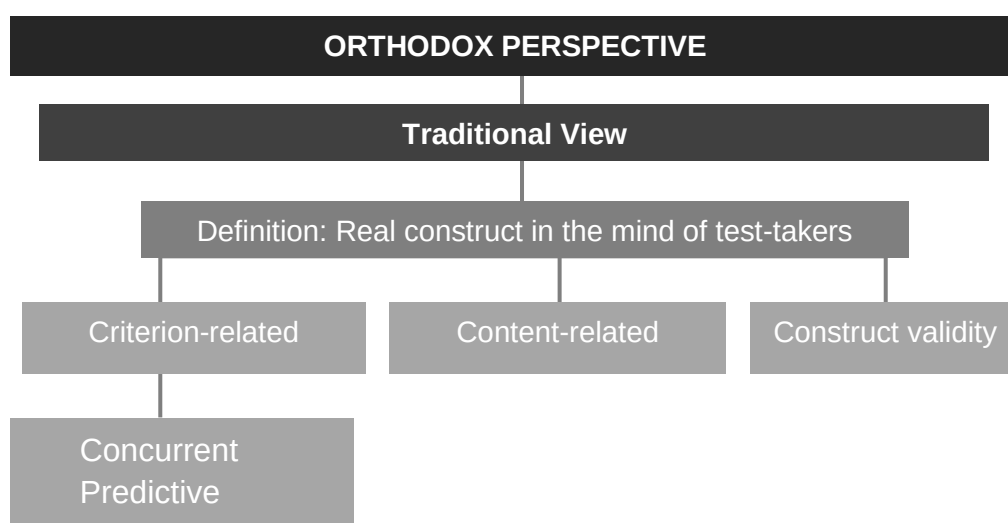


Figure 4.1: Diagrammatic representation of the orthodox perspective

The orthodox perspective is the traditional way of how validity was regarded in the beginning – from the early 1960s to the early 1980s. Proponents of the orthodox perspective of validity include, among others, Lado (1961), Spolsky (1975) and Oller (1979). The manner in which they define validity is comprehensively summarised by Van Der Walt and Steyn (2008:192) as referring to a real construct in the minds of test-takers (the ability or trait existing within the minds of test-takers for testing) as well as validity being an inherent characteristic of a language test. For instance, in his definition of validity, Lado (1961:321) would ask, “Does a test measure what it is supposed to measure?”. The answer would then be that, if it does, it is considered valid. One realises from this definition that validity was considered to be an inherent characteristic of a language test. He also considered validity to be one of the two important qualities of a test, of which the other was reliability. This is clearly demonstrated by one of his contemporaries, John Oller. Oller (1979:406) stated that the final standard by which the validity of language tests can be determined is based on the extent to which these tests reliably assess abilities of candidates to produce and process the discourse. This was based on the extent to which one test correlated

with another valid and reliable test. This is a mostly quantitative and a *posteriori* concept, meaning that this validation is performed after the test is taken. Criticism against this method of validation is that there could be a danger in placing one's faith in a measure that, in itself, may not be a measure of that upon which the construct is based. However, it should be noted that criterion-based validity like this can be split into concurrent and predictive validity. Concurrent validity is based on a criterion which is believed to be the ability which is being tested (Bachman, 1990:248). This could be test scores based on a certain measure of performance from older and longer tests taken at the same time, or teachers' rankings of students. Predictive validity, however, is based on a situation where the tester is just interested in the predictive power of the measuring instrument. It could be that all the tester needs, is to establish certain predictions about future performance on the basis of the results obtained. The correlational method of investigating validity as explained here, was seen as the most central method towards determining criterion-related validity (Chapelle, 1994:225).

From the definition of what validity means by Lado (1961), Oller (1979) and Spolsky (1975), as well as the summary by Van der Walt and Steyn (2008), the issue of 'construct' features prominently in the orthodox perspective. This definition highlights the existence of a construct in the minds of test-takers. This same construct was considered to be one of the three validities (namely, criterion-related, construct-related, and content-related validities) by proponents of the orthodox perspective. However, other proponents of this perspective, like Henning (1987), identified construct validity as one of the five validities which he investigated during the time period in which the orthodox perspective was popular. The current study will, however, not discuss other validities that were investigated by Henning, as those validities were not part of the prevailing perspective at the time. This study simply mentions construct-related validity at this juncture to entrench the prevalence of construct validity at the time when this validity type was prevalent. The important role played by a construct – in particular the importance of a common underlying construct between different tests – is of special importance in this study. Section 4.4 outlines how such a common construct was identified from the South African NCS for the purpose of developing the

TALA. In addition, Section 4.3 investigates how a test construct is validated according to the *a priori* validation process utilised for this study.

The preceding discussion supplied a definition of what criterion and construct validities entailed according to the orthodox perspective; and hence it is important also to supply information about what content-related validity entailed according to this perspective at this point. Content-related validity was used to refer to the content coverage of the task, in its basic form. Additionally, in its abstract form, content-related validity was considered as consisting of expert judgement about test content that showed results from empirical research consistent with theory-based expectations. Content analysis such as this, was presented as one of the methods used to investigate construct validity. Discussions on construct validity throughout this study will help to show the relationship between construct validity and content validity and how this is applied in the development of the TALA test.

Based on the above discussion, one can say that an orthodox perspective of validity focused on the so-called trio of validities (Chapelle, 1994:256). This discussion has also highlighted that criterion-related validity (which was part of the trio) was considered to be central to validation (Chapelle, 1994:255), as made up of concurrent and predictive validities. This discussion also entailed basic information on construct validity and how it relates to this study. The diagrammatic illustration at the beginning of this section represents in visual format a breakdown of this perspective.

4.2.2 The postmodernist perspective

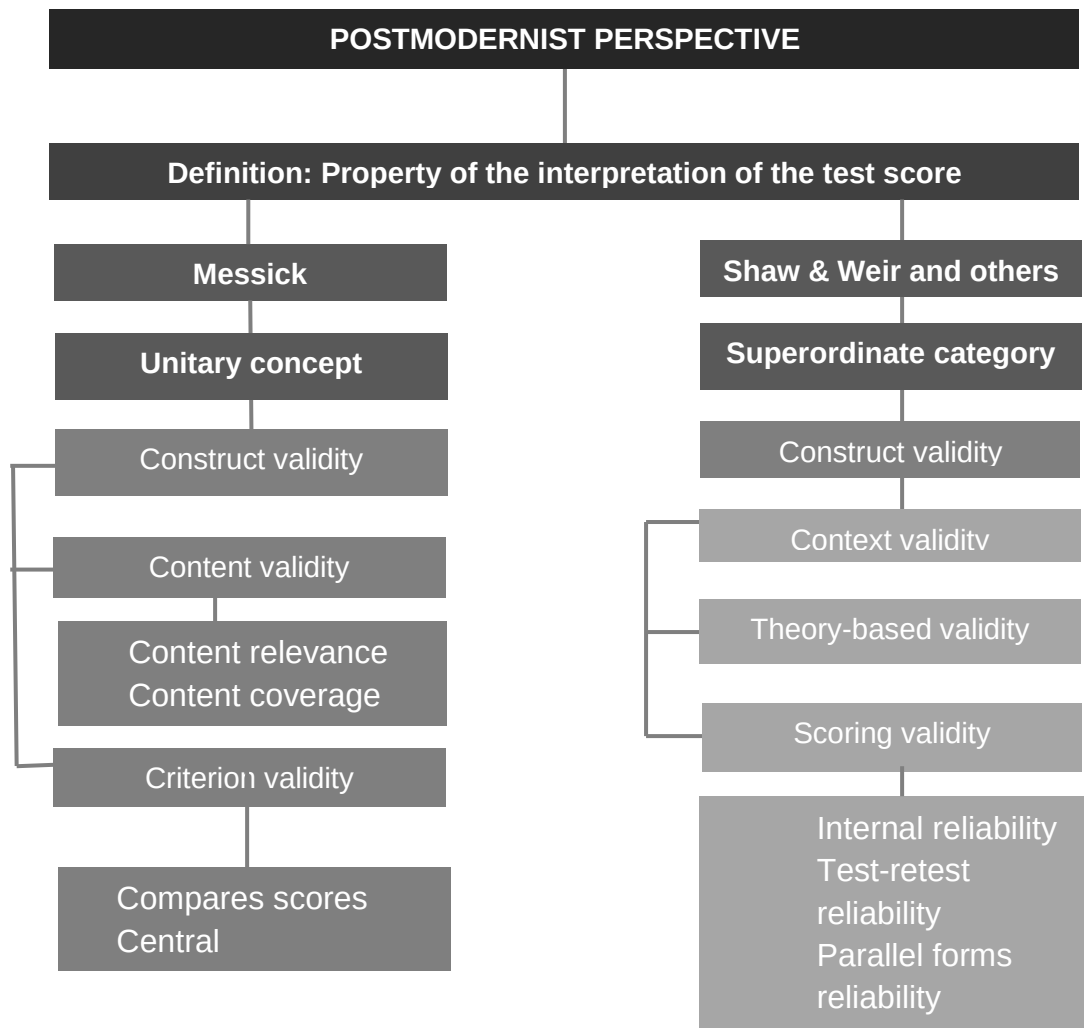


Figure 4.2: Diagrammatic representation of the postmodernist perspective

Based on the American psychologist, Samuel Messick's (1989) conception to incorporate test consequences and ethics – such as washback and social responsibility – into a theory of validity, a postmodernist view evolved. This new view of validity gave birth to a perspective that came to be regarded as postmodernist because it viewed aspects of validity differently from how it was viewed before. For this reason, a postmodernist perspective may be defined as a recent perspective that developed immediately after the orthodox perspective. It was conceptualised first by Samuel Messick, and was elaborated further by other scholars. From this perspective, validity is regarded as a property of the interpretation of test scores as highlighted in

Figure 4.2 above. This, according to Messick (1989:13), meant that validity is a judgement based on how far “empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores”. From this perspective, a new notion that introduced a unified model of validity emerged. It came to be called a unified model of validity because of the different forms of validity that were incorporated into construct validity. These different forms of validity that were incorporated into construct validity were content validity, criterion validity and predictive validity. Subsequently, construct validity came to be regarded as the central underlying component while content validity, criterion validity and predictive validity came to be regarded as mere aspects of it (Messick, 1980:1015, 1989:20). In this context, construct validity is defined as the extent to which a test accurately measures what it purports to measure, or the degree to which evidence and theory support the interpretation of test scores held by test users (Messick, 1990:8). Based on Messick’s definition of construct validity, it is important to examine how other aspects are defined by this researcher and how they are related to construct validity. It is, therefore, important to point out that the biggest difference between the orthodox and the postmodernist perspectives lay in the manner in which each perspective defined and explained each concept. Messick (1980:1017) defines content validity as “the relevance and representativeness of the task content used in test construction and does not refer to test scores”. He differentiates between content relevance and content coverage. By content relevance he meant specifications of the behavioural domain in question and the attendant specifications of the task or test domain. By content coverage he meant the specification procedures for sampling the domain in some representative fashion. In addition, he defines criterion validity as that type of validity where validity is “evaluated by comparing the test scores with one or more variables (called criteria) considered to provide a direct measure of the characteristic or behaviour in question” (Messick, 1990:11).

In addition to the above-mentioned unitary concept of validity as espoused by Messick, other scholars, such as Weir (2005a), added subcategories such as context validity, theory-based validity, scoring validity and external validity and referred to this unitary

concept as a superordinate category. Shaw and Weir (2007) went on to refer to the subcategories mentioned here as categories making up construct validity. Moreover, Shaw and Weir (2007:7) no longer considered reliability as a separate characteristic of a test, but an aspect of scoring validity. It is important at this stage to explain each of these subcategories of validity (as was done with each category of validity advocated by Messick) in order to illustrate how Shaw and Weir's subcategories differ in terms of definition and explanation from those of Messick. This part will also provide specific reasons as to why these subcategories were developed by researchers.

Context validity could be considered to refer to what used to be referred to as content validity traditionally (Taylor & Weir, 2005:152; Weir 2005a:19). It can be defined as a condition under which a task is carried out and is more inclusive and superordinate. It therefore includes discursal, social and cultural conditions as well as different linguistic parameters under which a task is performed (Taylor & Weir, 2005:152). Explaining the scope of context validity alone, Weir (2005b:152) asserts that "it is concerned with the extent to which the tasks chosen in a test are representative of the larger universe of which the test is assumed to be a sample".

The next subcategory of a unified concept of validity advocated by Weir (2005b), Shaw and Weir (2007) and Taylor and Weir (2005b) worth discussing is theory-based validity. Theory-based validity is sometimes described as construct validity (Taylor & Weir, 2005:237). One understands why they describe it as such when one considers the definition which these authors provide for this type of validity. They define it as "the adequacy with which the test items, tasks, topics, and language dialect represent the construct or theory or underlying trait that is measured in a test"(Taylor & Weir, 2005:237). The explanation of two prominent aspects of theory-based validity below should also shed more light on why this validity is also described as construct validity.

According to Weir (2005a:17), more attention was given to gathering empirical *a posteriori* (after) evidence which was purely statistical than *a priori* (before) evidence which was more theoretical, before Messick. The focus on *a posteriori* evidence "lent itself easily to discrete elements of the target language (lexical or grammatical items)"

and little to “*a priori* deliberation on the match between theory and test” (more information about *a priori* and *a posteriori* validation is given under Section 4.3)(Weir, 2005a:17). For this reason, construct validity was seen from a purely statistical perspective.

The superordinate, unitary concept of validity also includes another important subcategory that should be explained here. This refers to scoring validity. Scoring validity is what used to be referred to as reliability by many authors before the times of Shaw and Weir (2007). It was defined as the “degree to which scores in a test display consistency of measurement” (Bachman & Palmer, 1996:19). As mentioned earlier, after the times of Shaw and Weir, reliability came to be regarded as no longer a separate characteristic of a test, but an aspect of validity. Scoring validity came to be considered as a superordinate term for all aspects relating to reliability (Weir, 2005a:14). However, for obvious reasons, much of the traditional literature still explains issues such as internal consistency estimates and parallel forms of reliability of tests as just reliability. Some of the tests dealt with in this study (mostly from the South African DBE) are criticised on their failure to achieve reliability. This type of labelling is retained in order to avoid sacrificing certain aspects of validity for the ordinary pursuit of reliability as so many traditional authors did (Weir, 2005a:7). However, this study will generally assume the point of view where reliability is considered as scoring validity according to the contemporary view.

It is important to recognise that there are different types of scoring validity when dealing with the issue of test measurement. Traditionally, these types of validity were labelled test-retest reliability, parallel forms reliability, internal consistency, and marker reliability. Test-retest reliability is explained as a situation where the same test is given to the same group of learners twice to measure the consistency of achievement. “The scores obtained by the same persons on the two test administrations are correlated to give a reliability coefficient” (Weir, 2005a:25). In this regard, the reliability coefficient can range between -1 and +1 where the number 0 indicates a total lack of reliability and +1 indicates complete or perfect reliability (Weir, 2005a:25). This type of reliability can be problematic, however. The problem is caused by the period of time that elapses

between different test administrations. Some authors suggest that the time taken between test administrations may not be more than six months and that the time between administrations be constantly monitored as it may affect the test score outcome (Weir, 2005a: 25). The gap in terms of time between tests can influence the outcome between tests and correlations can decrease progressively as the interval increases. In order to overcome the gap, according to Weir (2005a:25), once reasons for poor performance by learners because of time lapse between sessions have been established, it might be worked out what the correct answers should be from answers given by learners and many interventions to try and improve their responses might not be necessary.

With regard to parallel forms reliability, the reliability coefficient is calculated through the usage of different tests but tests reflecting similar specifications such as item length and breadth, degree of topic familiarity and difficulty level. The results attained by learners for the first test would be statistically compared with the results attained by them for the second test. The correlation of these tests results will indicate the parallel forms reliability of these two tests. This would be an estimate of the extent to which the two forms awarded the same marks and would indicate how much error of variance had resulted due to content sampling of the two forms. This testing approach has been used by many testing bodies to level criticism against tests and examinations not reflecting parallelness of form between tests, but this does not render this reliability type free of its own practical shortcomings, some of which are similar to those of test-retest reliability. An example of the obvious practical shortcomings of test-retest and parallel forms reliability is similarity of tests. The same tests are offered to the same learners after a certain time lapse in a test-retest format, while in a parallel forms format the two tests need to be similar in many respects and be administered to the same learners, a situation which renders each test generating a certain measure of experience or wash-back to learners.

Based on the problems that might be encountered with the test-retest and parallel forms reliability, other measures such as internal consistency measures might be necessary. The internal consistency reliability measure is based on Classical Test

Theory (CTT) or Item Response Theory (IRT), including Rasch Analysis. CTT generates a single estimate for all candidates while IRT provides an estimate for the score of the individual candidate. These procedures help to identify test items that are not consistent in the manner in which they test abilities and they also help to identify items which perform poorly in discriminating between candidates that differ in ability. Based on the application of CTT, determining the internal consistency of a test is a standard procedure. This implies the level of agreement between the results of one test with itself or with another test. Van der Walt and Steyn (2008:195) explain the notion of internal consistency of a test through the illustration of a test with dichotomously placed subsections. They indicate that “if a test consists of a number of sections with dichotomously scored items, the internal consistency for each section and for the whole test can be determined by calculating the Cronbach alpha coefficients” (Van der Walt & Steyn, 2008:195). Where learner performance on one half of a test’s items is compared with the learner’s performance on the other half of the items, this reliability measure is referred to as split-half reliability. A requirement for this reliability is that items should test the same ability and be dichotomous. If these items are ranked in order of difficulty, one half should be 1, 3, 5, 7, and another 2, 4, 6, 8. Theoretically, the performance of the student should be the same for each half. According to Weir (2005a:29), the two scores of the two halves can be correlated. Where the answer is 1,0 in correlation, it indicates a perfect match and 0,7 indicates about fifty percent agreement between two sets of scores. According to Weir, a reliability estimate of 0,8 is normally considered a minimum acceptable level but the expectation normally is that it would exceed 0,9 in tests of importance or high stakes tests. Van der Walt and Steyn (2008:196) go on to indicate that the alphas of sections of a test have an impact on the construct validity of a test. The construct validity of a test is questionable if the alpha coefficients of sub-tests deviate from the norm.

Van der Walt and Steyn (2008) go on to state that the alpha coefficient of a test is a function of the number of items on which it is based. Therefore, this coefficient increases with the number of items. According to them, the accepted norm for the alpha coefficient is 0,8, but the reliability coefficient between different sections of tests

should be lower than this measure to show that different sections measure different aspects. For this, they quote Alderson *et al.* (2005:184) and indicate that the alpha coefficient between different sections of a test should be 0,3–0,5 (Van der Walt & Steyn, 2008:196).

The difference between CTT and IRT, according to DuPlessis, (2017:83) is based on the fact that CTT focuses mainly on “descriptive statistics, determining the standard error of measurement, reliability estimates and item analysis”, while IRT, according to her, “is suitable only for the analysis of larger groups” and going beyond the “descriptive and inferential statistics to probabilistic models that allow a further measure and level of generalisation” (Du Plessis, 2017:83).

It has been explained above that in dealing with parallel forms of reliability, error variance may emanate from test-retest or parallel forms of reliability. However, a third instance of error variance might emanate from various scores involved in the process. Different assessors marking tests may want to evaluate their marking against each other or within themselves. They may, therefore, want to check if they agree about the scores which they give to the candidates or the manner in which they rank their candidates. When markers deal with samples of writing or speaking, they therefore want to estimate the consistency of their marking. This could be done in two different ways, viz, the consistency of their own individual marking termed intra-rater reliability and the consistency between themselves as markers termed inter-rater reliability. In order to mediate for an objective inter-rater reliability in a more empirical way in writing and speaking, markers currently apply more sophisticated Item Response Theory (IRT) statistical models. A Multifaceted Rasch (MFR) analysis is a good example of what may be applied for this purpose. This model has the ability to identify an inconsistent marker over time and in comparison with other markers.

All the above-mentioned *a posteriori* scoring validity measures are based on certain formulae and calculations based on descriptive statistics. It does, however, appear that in terms of contemporary applications, people do not use these formulae frequently, they simply use programs such as *Iteman 4.3* and *Tiaplus* to do such

calculations. Be that as it may, it is important to understand the computation of how these programs arrive at their scores. This will be demonstrated in Section 4.2.2.1. Descriptive statistics also help with the interpretation of other statistics in order to generate validity arguments. It does this through the usage of histograms, frequency tables, measures of dispersion and measures of general tendencies. These will be further explained in the following section.

4.2.2.1 Descriptive statistics

The analysis of specific test items includes issues such as the calculation of facility and discrimination values, as well as distractor efficiency estimates when using multiple-choice tests. It should be noted at this point, that the TALA is one such multiple-choice type test. Normally, specific weaknesses of test items can be identified by using the calculations referred to above. According to Brown (2012:327), facility values show the distribution of correct and incorrect answers in percentages for the population of test takers, i.e. how 'easy' or 'difficult' an item is. Brown (2012) asserts that according to CTT, facility values that are recommended for retaining items are those between 0,30 and 0,70, neither being too easy, nor too difficult.

A test item can have either a high or a low discrimination index. This implies that a test item could distinguish between a weak and a strong test taker based on the discrimination index of such a test item. An item discrimination value indicates the performance of a test taker on a particular test item as related to other items in the whole test.

The Pearson point-biserial correlation is another important factor in reliability measurement in descriptive statistics. It normally indicates the differentiating strength of a test item and normally ranges between -1,0 and 1,0. A negative value indicates that test takers are not responding appropriately to the item. It can mean that strong test takers are answering a specific item incorrectly while weaker ones are answering correctly. When the point-biserial value is 0, it means that the item is failing to differentiate between strong and weak test takers. This is what has previously been

referred to as measurement error as was mentioned in the introduction of this section regarding various factors that may cause measurement errors. In order to calculate the true score of a test taker, a certain measure of error must be considered as, according to Fulcher (2010:4), no language test can be fully reliable because language ability cannot be observed in a consistent and predictable manner. This is why the formula to calculate a test score (X), includes a true score (T) as well as the error (E). That formula is thus as follows:

$$X = T + E \text{ (Fulcher, 2010:47)}$$

In order to calculate the effects of measurement error, a formula that includes reliability estimates, the standard deviation and the reliability coefficient are required. The combination of these variables produces the following standard error of measurement (SEM) formula: $SEM = SD\sqrt{1-R}$. The calculation of the SEM is important because, according to Du Plessis (2017:86), “through the calculation of the SEM, a confidence interval can be generated in respect of any score to indicate how much this score may fall below or above the observed score”. The SEM was also used to estimate reliability of a test before the advent of computers by comparing test results of the same test over a period of time, but currently reliability is estimated only by the internal consistency estimates of a test.

According to Brown (2012:326), the following formula is used to measure the reliability of a test based on CTT:

$$=Reliability \frac{var_{true\ score}}{var_{true\ score} + var_{error}} = \frac{var_{true\ score}}{var_{total}}$$

This formula is based on the notion that true scores do not have errors of measurement and that variations to true scores can occur. These variations to true scores are termed “true score variance”. The above-mentioned formula implies that true score variance divided by total variance in a test gives reliability.

The formula used to measure the reliability of a test operates on the notion called the split-half reliability as indicated in Section 4.2.2. This formula divides items of a test

into two components based on whether the item is an even number or an odd number. This formula is used to obtain the correlation coefficient between these two separate groups of items. It is thus used to generate the estimate of the two separate halves of the test. The reliability estimate of these two halves should be lower than that of the full test. After the half test correlation has been obtained, it should be adjusted according to the Spearman-Brown prophecy formula which will produce a correlation for the two halves of the test. Fulcher (2010: 51) provides the following formula for this operation:

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

However, other formulae such as the K-R20, computes different variables to produce this outcome. The K-R20 uses the number of items (k), standard deviation (SD), the proportion of examinees who provided the correct answer per item (p), and the proportion of examinees who provided the incorrect answer (q) to produce the required correlation between the two halves of the test. Its formula, according to Brown (2012:329), is as follows:

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

The Cronbach alpha, however, has more advantages than the two above-mentioned formulae in terms of its flexibility when it comes to weighted items and is highly recommended for determining the reliability of a test.

The following section will demonstrate how descriptive statistics can assist with the interpretation of other statistics in order to generate validity arguments. As pointed out earlier in Section 4.2.2, it does so through the use of histograms, frequency tables, measures of dispersion, and measures of general tendencies.

Histograms and dispersion tables in descriptive statistics indicate the spread of test scores and range and are applied extensively in this study to try and illustrate differences between the scores of test-takers. In addition, it is by means of these

aspects of descriptive statistics that the anomaly between the performance of different HLs in the South African education system was identified. According to Fulcher (2010:40), a distribution curve is used to analyse norm-referenced language papers such as those for HLs. A norm referenced paper is a paper in which individual test takers' scores are compared relative to those of other test takers. A normal distribution curve should be able to produce a bell-shaped curve and the majority of scores – about 68% thereof – should belong close to the mean of the curve while each of the 34% thereof should belong under the mean and over respectively. According to Du Plessis (2017:85), it was on the basis of an interpretation of the mean obtained per HL over a four-year period – between 2011 to 2013 – that Umalusi realised the HL papers may not be comparable in terms of standard or construct. Note the averages of the Grade 12 NSC examination for 2011 in Table 4.1 adapted from the Daily Press (Rademeyer, 2012:7).

Table 4.1: “Results of the Grade 12 examination of 2011” (Weideman *et al.*, 2017:3)

Subject: home language	Average mark (%)	Subject: First Additional Language	Average mark (%)
English	55,7	English	47,2
Afrikaans	59,3	Afrikaans	52,6
Isixhosa	63,7	Isixhosa	65,0
Isizulu	67,2	Isizulu	75,2
Pedi	59,8		
Sesotho	60,3		

The averages (means) displayed in Table 4.1 enabled Weideman *et al.* (2017:3) to come to the conclusion that the exit-level assessments for some South African

languages barely discriminate among learners of different abilities, while others do. As can be seen from Table 4.1, the average examination scores of African HLs are higher than those of English and Afrikaans and the average scores for African first additional languages (FALs) are higher than those of English and Afrikaans FALs. This gives the impression that English and Afrikaans HLs as well as English and Afrikaans FAL papers are more difficult than those of African HLs and African FALs respectively. Therefore, means such as these help researchers to determine trends and tendencies.

Based on this analysis, one realises the importance of a mean or average score in descriptive statistics. However, the role of a mean in descriptive statistics does not end there. According to Bachman (2004:63), the standard deviation refers to the extent to which test scores tend to vary or deviate from the mean on average. This implies that if the mean score for all averages recorded in the table is calculated and each score is compared with that mean, it would produce the standard deviation. For example, in Table 4.1, the average score for all the stated averages is 61% and the standard deviation for Sesotho HL would be 0,7. Therefore, the mean also features prominently in the calculation of the standard deviation in dispersion measurement and in the determination of the standard error of measurement.

The last important aspect to consider according to the superordinate concept of validity, is external validity. This is considered when enough evidence on the context, underlying theories and scoring has been generated. Determining the external validity of a test means going outside of the test and evaluating its impact and its relation to other measures of the same ability. This aspect of validity was originally conceptualised by the proponents of the orthodox school. This external validity consideration also includes consequential validity and was not discussed in the introduction under the orthodox perspective. Consequential validity originated with modernist perspectives to validity and is part of Messick's conception of validity. It deals with issues of bias and potential sources of bias. Bachman (1990) identifies three possible sources of bias: cultural background, background knowledge, cognitive characteristics. These issues should be statistically measured after the test was administered. They are electronically recorded and compared to test scores to check

77

their impact on the score of the candidates. The issue of bias in testing is clearly demonstrated through the process of validation which will be discussed comprehensively in Section 4.3.

As can be seen from the discussion above, on a postmodernist perspective of validity, the view of a unitary concept of validity, which later evolved into a superordinate view, was maintained throughout the 1990s by experts such as Bachman (1990), Chapelle (1994), Cumming and Berwick (1996) and Kunnan (1997) and is clearly outlined diagrammatically in Figure 4.2. However, this notion of a unitary and superordinate concept of validity has also been criticized when testing experts tried to explain and elaborate on it. It was, for example, felt that as much as this concept is influential, it is also complex and demanding and difficult to operationalise. For this reason, Bachman and Palmer (1996:17) introduced the notion that 'usefulness' was the most important quality in language testing, and a contemporary perspective of validity was born.

4.2.3 The contemporary perspective

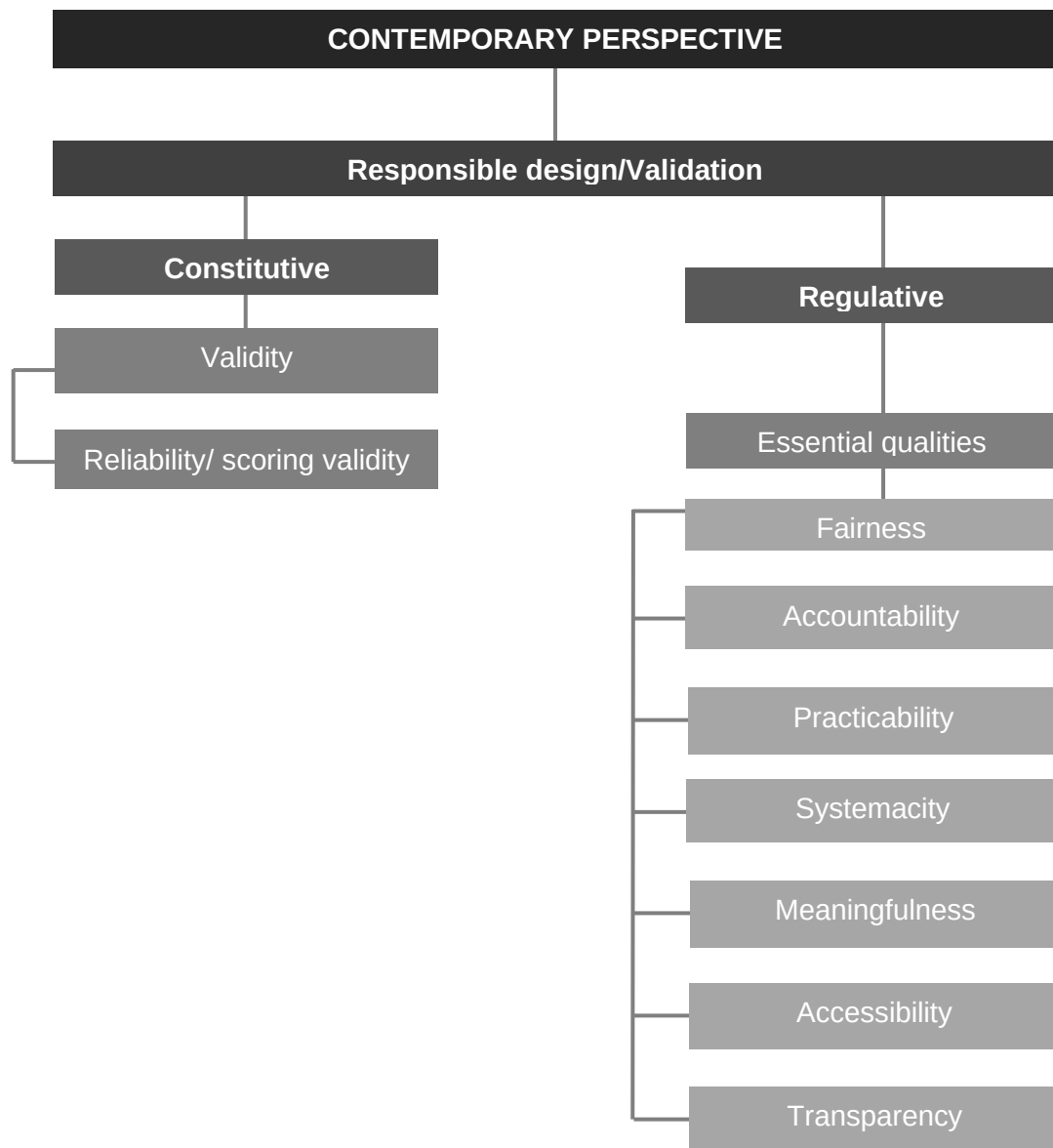


Figure 4.3: Diagrammatic representation of the contemporary perspective

Under the so-called contemporary perspective of validity, scholars such as Bachman and Palmer (2010), Fulcher (2015), Fulcher and Davidson (2012), Green (2014), Kunnan (2010), and Rambirich (2012) introduced the notions of the 'primacy of fairness', accountability and practicability, as the most important concepts in language testing. While Bachman and Palmer (2010) and Kunnan (2010), for example, subscribed to a unitary concept of validity, their attempt to clarify it and the accompanying difficulty in doing so, inadvertently fostered a new paradigm. They advocated that additional aspects such as 'primacy of fairness', accountability and practicability, are essential and necessary to be incorporated into validation. These modifications led to a situation where the superordinate notion of construct validity was contested. One such contestation is a view by Rambiritch (2012), that Messick's superordinate notion of construct validity relies heavily on empirical data and statistical procedure which may not be adequate to account for all aspects involved in measuring language ability. Rather than placing too much emphasis on the superordinate notion of construct validity, researchers such as Bachman and Palmer (1996), Green (2014), Rambiritch (2012), Van Dyk (2010), Weideman (2009) and Weir (2005b) just to name a few, emphasise essential qualities of tests that are necessary for validation. Some of these scholars subscribe to the notion of validation as 'responsible design'. This responsible design refers to the ability of a test developer to incorporate the concepts of fairness, accountability and practicability in the process of collecting useful evidence for the validation of the test. In their bid to re-interpret and extend Messick's validation concept, different scholars use distinct angles to incorporate the notion of responsible design in their validation methods, which gave rise to the 'contemporary' perspective on validity. Bachman and Palmer (1996:17), for instance, while acknowledging the concept of validity in testing, emphasise the issue of the 'usefulness of the test' as the most important consideration in language test design and development. They subsequently identify four qualities as important to the useful assessment of a test, namely, practicability, reliability, validity and beneficial consequences. This implies

that the essential quality for a test, according to Bachman and Palmer (1996:17), is "the extent to which a test is useful to the tester and the test-taker".

Another scholar who subscribes to the contemporary perspective of validation is Weideman (2012:8). He does not conform to the issue of subsuming everything under a unitary notion nor under a superordinate concept of validity, but focuses on the concept of responsible design. He draws this concept from his paradigm of applied linguistics which he views as a 'discipline of design'. He supports a framework for language testing which is based on two main tenets of an analytical and theoretical nature, which he refers to as constitutive and regulative conditions. He concludes that rather than conflating everything under one unitary notion, they should be viewed separately as having relevance and being interrelated. In his article, "Validation and validity beyond Messick" (Weideman, 2012:8) that focuses on conditions for responsible design, Weideman identifies two studies, one by Van Dyk (2010) which involves the Afrikaans 'Toets van Akademiese Geletterheidsvlakke' (TAG) and the other by Rambiritch (2012) that focuses on the 'Test of Academic Literacy for Post Graduate Students' (TALPS), where a process of responsible design has been followed in the development of these two tests. Weideman (2012) emphasises that although Van Dyk's study (2010) was initially conceived on the basis of constitutive requirements as a validation study, it was later reconceptualised to include regulative requirements for responsible design such as systematicity and meaningfulness of results. Similarly, Rambirich's study (2012), which was also initially based on constitutive requirements, was later reconceptualised to include the regulative requirements of responsible design such as accessibility, transparency, and accountability. These two studies serve as tangible examples that subscribe to a contemporary perspective of validation.

The group of researchers who could be considered to belong to the contemporary scholars also includes Chapelle, Enright, and Jamieson (Chapelle *et al.*, 2010). The reason for their inclusion here is the radical stance they adopted when they evaluated the difference between the argument-based approach towards validity as presented by Kane (2006) and which was described in the 1999 AERA/APA/NCME 'Standards'

for Educational and Psychological Testing. In this evaluation, Chapelle *et al.* (2010) examine the difference between these two approaches according to four points of comparison, namely: “framing the intended score interpretation, outlining the essential research, structuring research results into a validity argument, and challenging the validity argument” (Chapelle, 2010: 11).

Under the point, ‘detailing the framing of the intended score interpretation’, Chapelle *et al.* (2010:3) conclude that, since the ‘standards’ frame the intended interpretation of the test as a construct which can be defined from many different perspectives and strong construct theory could not generate clear hypotheses about test performance relative to other constructs and behaviours, the construct itself was not a good basis for subsequent research. This argument disqualifies construct as a strong basis upon which validity can be tested. By saying so, they opened a different path for the use of an argument-based approach to validity arguments.

Regarding the point, ‘outlining the essential research’, Chapelle *et al.* (2010:5) state that generally, either a construct or interpretive argument is a starting point for planning research to be used in a validation argument. However, they go on to state that it is almost impossible to define a construct for activities such as ‘reading comprehension’ and ‘speaking ability’. Therefore, the argument-based approach becomes necessary for activities such as these. Additional reasons that support an argument-based approach include Kane’s (2006) assertion that evidence that supports test scores should be generated while ‘standards’ assert that test developers and researchers should gather ‘kinds of evidence’ that are needed to evaluate the ‘intended interpretation’ of test scores (Chapelle *et al.*, 2010:6). While trying to generate propositions in line with evidence gathering as proposed by ‘standards’, (Chapelle *et al.*, 2010:6) Chapelle *et al.* realised that the propositions they were generating could not be exhausted and were based on research that had been finalised and not on scores that were newly produced. For this reason, Chapelle *et al.* (2010) realised that in the absence of guidelines about clear propositions they were generating, they were

influenced by the validation research that had been completed, and were therefore unlikely to help identify areas where more research was needed.

Kane's (2006) argument-based approach also includes 'inferences' and their 'assumptions'. These 'inferences' and 'assumptions' are necessitated by the need to identify framing statements to which 'inferences' and 'assumptions' could be tied when formulating interpretive arguments that would ultimately lead to validity arguments. A given score in the assessment of speaking for the Test of English as a Foreign Language (TOEFL), for instance, should be supported by a warrant such as hesitation and mispronunciation. A warrant, according to Chapelle *et al.* (2010:6), is a law, a generally held principle, rule of thumb, or established procedure and 'assumptions' would be, for example, that the rubric for scoring the responses was appropriate for providing the relevant evidence for ability. Based on concepts such as warrants and assumptions illustrated here, an 'inference' for a score that would be given is made. This string of reasons that is connected to the score is referred to as an interpretive argument. It should be noted, however, that Kane includes even more 'inferences' to this interpretive argument. He actually gives six different 'inferences' before one can arrive at a concluding interpretive argument (Chapelle *et al.*, 2010:8). It must further be noted here that, contrary to an infinite number of propositions required by 'standards', Kane's 'inferences' are bound and are achievable in order to arrive at a segment of an interpretive argument that can clearly lead to an argument-based approach to validity.

With regard to the matter of 'structuring research results into a validity argument', Chapelle *et al.* (2010: 9) explore the issue of listing types of evidence. They explain that 'standards' recommend that research be identified where the validity evidence remains a listing of research results which researchers need to integrate in a narrative and which others need to evaluate in view of their own experience. In addition to this, they state that, as far as the TOEFL is concerned, a taxonomy consisting of types of research that have been conducted is maintained. They go on to state that such a list could also be used to identify areas where more research is needed. However, they conclude by stating that a "taxonomy is not an argument, and in working with a

taxonomy one is not prompted to look at strength of evidence or to organise it in a way that presents a validity argument” (Chapelle, 2010:9).

Kane, according to Chapelle *et al.* (2010:9), depict the validity argument as a step-by-step movement across inferences to a conclusion about test score use as partly stated already. “The interpretive argument lays out the intended inferences and warrants” (Chapelle, 2010:9). This implies that “the process of building a validity argument requires the researcher to examine the assumptions underlying each of the warrants in order to focus on obtaining the relevant evidence” (Chapelle, 2010:9). Kane, Chapelle, Enright, and Jamieson also say that this process provides the evidence “needed to move to the conclusion that the observed score is appropriate” (Chapelle, 2010:9). Conversely, the generalisation which can be defined as the generalisation inference is supported by evidence which consists of generalisability and reliability studies together with the scaling and equating studies based on the observed score. In addition to this, there is also what can be termed the explanation inference which is the concluding part. “The explanation inference links the expected score to the construct of the test” (Chapelle, 2010:9). The explanation inference is supported by several studies which are typically associated with construct validation. These studies might be studies such as discourse and cognitive processes studies, task characteristic and item difficulty studies, concurrent correlational studies, analysis of reliability and factor analysis and comparison studies of group differences. Finally, as stated by Chapelle *et al.* (2010:10), use is made of the so-called utilisation “inference which links the target score to the conclusion” that the score is useful for aiding admissions, or placement, and guiding instruction of the language (if the test is meant for these purposes). This process is based on the TOEFL validity argument specifically and may differ depending on the type of test.

According to Chapelle *et al.* (2010:11), the issue of ‘challenging a validity argument’ “is described in terms of counterevidence for propositions”, whereas Kane’s approach refers to refuting the argument. The outcome of this comparison, according to Chapelle *et al.* (2010:11-12), revealed that the approach suggested by “‘standards’ lacked the structure required to move forward with decisiveness and the reliance on

the test construct as a starting point". The 'standards' were also problematic not because the test was solely focused on performance without interest in a construct, but "because the construct was extremely complex" (Chapelle, 2010: 12). However, the argument-based approach to validity according to Chapelle, provides "a place for the construct, without relying on it as a starting point. It also provides a means of specifying the multiple inferences underlying score interpretation and use, and thereby remove the burden that might otherwise be placed on an imprecise theoretical construct"(Chapelle, 2010:12).

The criticism outlined above culminated in a framework for validity that is explained by Chapelle (2011) in her article: "Validity argument for language assessment: the framework is simple". This framework, which was proposed by Chapelle *et al.* (2010), outlines the following important points identified from Kane's argument about validity.

First, it explains an interpretive argument to state the meaning which has many layers from test scores (Chapelle, 2011:19). Second, it defines the most necessary validation research by a process that is systematically carried out for making inferences in the argument that is imperative instead of going through a list of types of evidence of validity (Chapelle, 2011:19). Third, research results are presented in an integrative manner in order to show how they support the validity argument. This interpretive argument with supporting evidence is referred to as a validity argument (Chapelle, 2011:19). Fourth, the manner in which the interpretive argument is put together helps to interrogate, make it weak, make it less plausible or reject the research through criticisms (Chapelle, 2011:19-20).

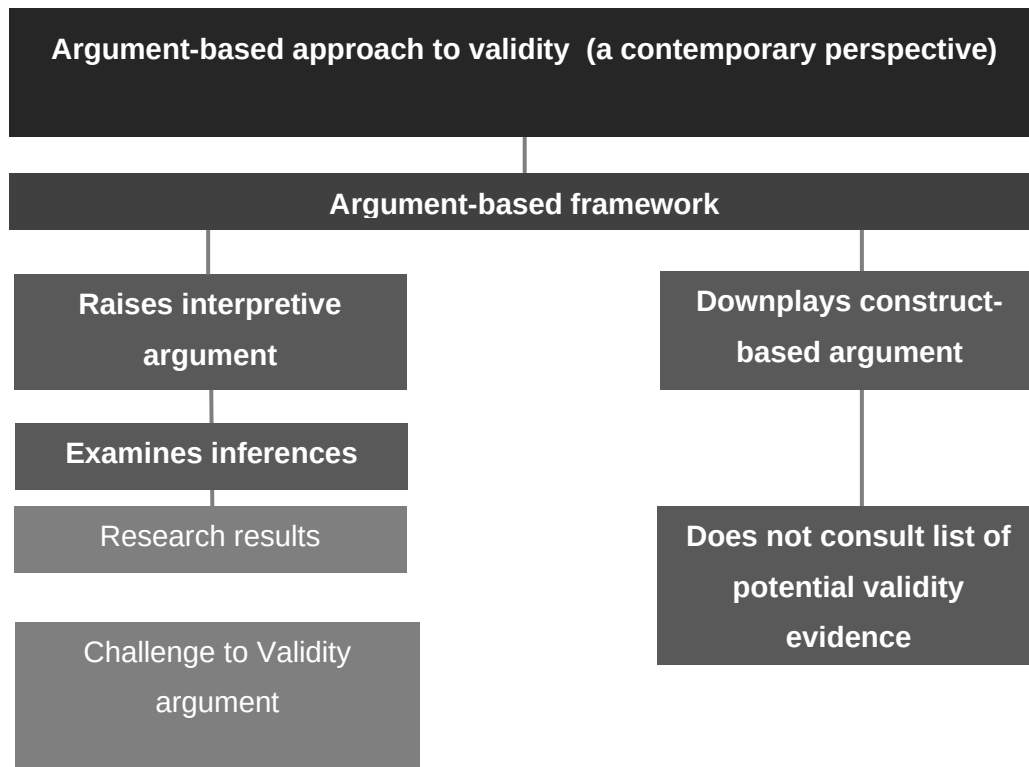


Figure 4.4: Diagrammatic representation of the argument-based approach to validity (a contemporary perspective)

For this study, an integrated approach which follows the contemporary perspective towards determining the validity and scoring validity (reliability) of the TALA was followed. Certain essential qualities of responsible design used contemporarily in Figures 4.3 and 4.4 for the validation of a test was used to determine the validity of the TALA.

4.3 The process of validation and various elements of evidence

Validity and validation are closely related concepts. While validity is an abstract concept as can be seen from the definition stated in Section 4.2.1, validation is the

process of operationalising it. Validation takes place within two different stages, namely, the *a priori* and *a posteriori* stages. The *a priori* stage is where evidence is collected before the test is administered and the *a posteriori* stage is where evidence is collected after the test has been administered, as explained earlier. The *a priori* stage is a part of validation that, according to Bachman and Palmer (1996:26) and Weir (2005:16-19), should be delineated right at the beginning of test or examination development. Weir (2005a:16-19) refers to this component of validation as encompassing both theory-based validity and context-based validity. The definitions of theory-based validity and context-based validity were provided in Section 4.2.2. Part of the *a priori* stage of validation involves the justification of theoretically assigning a particular construct to a test or examination and focuses on the validity of using a test for a particular purpose. It serves to outline the abilities which should be assessed in a test and how other abilities termed ‘construct irrelevant features’ may be grouped together into contexts that are not used when testing is done and are considered as irrelevant variables, e.g. method effect. The abilities to be assessed should be based on theoretically accepted principles of test design. At this stage of validation, enough tasks should be selected that are suitable for measuring the specified abilities and knowledge. Section 4.4 presents an analysis of the perceived construct for Grade 12 examination papers. What is clear from the discussion is that Grade 12 examination papers have to assess several important aspects in line with the NCS-termed Curriculum and Assessment Policy Statement (CAPS).

According to Van Dyk and Weideman (2004a:17-18), part of the *a priori* validation is that test designers have to evaluate which among different task types (and the items they contain) are best suited for the type of tests they need to create. They might discover that some of the items for tasks are not appropriate for the test based on certain requirements after piloting a test. They identify the following categories of item suitability when creating test items:

- Items that are accepted because they are suitable to the construct of the test.

- Items that may not be accepted because they are not productive and cannot be aligned to the plan.
- Items that are agreeable and are properly aligned to the construct and are highly productive.
- Items that are not ideal although they display a certain level of potentiality to be productive but are not quite aligned to the framework.

Making use of the categories of suitability described above, a test piloting phase enables one to refine a test by discarding test items that are deemed unsuitable. Therefore, not all the test items that are initially selected for inclusion may be retained for the final version of a test. Based on this, a test requires consistency of content. As part of a *priori* validation, test administration is also important. Tests should be administered under the same conditions to different groups of candidates with clearly defined and fixed methods of scoring (Fulcher, 2010:5). The TALA that forms part of this study is meant to feature as a component in each of the eleven HLs of the South African DBE. The considerations above would, therefore, need to form an important part of the *a priori* validation of this test.

With regard to *posteriori* validation, the following additional aspects are important to consider. A *posteriori* validation takes place over two stages viz., the test results and the evidence they produce of the ability of the candidates, and the empirical validation of the procedures by which judgements for the test were arrived upon (Van der Walt & Steyn, 2008:194).

The second stage of a *posteriori* validation is practically operationalised through the collection of evidence from different sources that relate to the test that was administered. This includes, among others, evidence of construct validity, content validity, criterion validity, and reliability indexes. This evidence also includes social aspects such as test-taker feedback, test consequences, test ethics, social responsibility, washback and the impact of test scores. It should, however, be noted that before a validation study is conducted, certain hypotheses about the test should

be formulated. These hypotheses are hypotheses such as: the test is based on a valid construct; its scores are reliable; it predicts future performance accurately; etc.

For a test such as the TALA which has similar properties to the TAG and TALL with regard to task types, a largely similar validation process to the one that was carried out by Van der Walt and Steyn (2008) for the TAG and TALL could provide important guiding indicators for this study. It is advantageous to do so because these tests have been researched extensively to include pertinent issues discussed above and have been developed based on a comprehensive blue print for further development. Therefore, the validation argument for the current study would involve the collection of *a posteriori* evidence making use of CTT, IRT and a factor analysis of test scores. It will also include evidence from test-taker feedback and test administration as social aspects that may influence the validity of test scores. It should, however, be remembered that making use of evidence from these sources belong to the postmodernist perspective explained in Section 4.2.2. It should be noted in this regard that the argument-based perspective still makes use of evidence from these sources, it uses such sources to empower argument-based argumentation to construct a strong validity argument.

4.4 Theoretically analysing the usage of the underlying common construct for TALA

Studies already concluded in the wider project of which the current study forms part, identified one of the causes of inequivalence between Grade 12 examination papers in the DBE in South Africa as the lack of a common understanding of the National Curriculum statement's underlying construct for assessment. Umalusi (2011) had already noted the contention which existed long before the introduction of the NSC in its report: "Making educational judgements: Reflections on judging standards of intended and examined curricular". This report emphasises the need to introduce equivalence between examination papers. Thus, after being commissioned by Umalusi to investigate this problem, the project team decided to analyse the New Curriculum Statement (NCS) titled the Curriculum and Assessment Policy Statement

(CAPS) (Grades 10-12) for a better understanding of the common construct between examination papers. This section will, therefore, explore sources of the DBE such as the NCS and CAPS to outline the rationale upon which the common underlying construct for CAPS is based. It will also supply information from sources of the project team to critically evaluate the basis upon which this underlying construct was conceptualised, in order to make an argument about the theoretical defensibility of the construct (Patterson & Weideman, 2013).

Since construct validity is an important principle of language testing that forms part of the paradigms upon which equivalence of tests rests, it is important to explore the basis from which the underlying construct of CAPS emanates. It is also one of the types of validity which this study will measure in order to determine equivalence between various Grade 12 HLs. From the analysis that was conducted by the project team, it is clear that the common underlying construct of CAPS emanates from the differentiated communicative competence of the 1970s as advocated by Habermas (1970); Hymes (1971); and Halliday (1978). This becomes evident if one considers that it pins its teaching approaches upon being text based and communicative (DBE, 2011:11) and upon the notion that language structures be learned in an integrated way.

In addition to the above, the team noted that the curriculum identified as one major objective, the enablement of learners to operate in a number of diverse contexts; namely, participating as responsible citizens in a free country and making the transition from educational institutions to the workplace. Moreover, CAPS underwrites principles of “high knowledge and high skills” as minimum standards to be achievable in home language subjects (Du Plessis *et al.*, 2013). These principles apply at different levels at which a language is offered, i.e., at HL level or additional language (AL) level. However, these principles do not refer to the levels of language offerings, like additional or home language levels, but the level of competency to be achieved at a specific level, which makes the principle of ‘high knowledge and high skill’ applicable to additional languages as well. CAPS also distinguishes the social from the educational levels of language proficiency as important levels at which the HL

90

operates. The social level is described as ‘the mastery of basic interpersonal communication skills required in social situations’ and the educational level as ‘cognitive skills essential for learning across the curriculum, including literary, aesthetic and imaginative ability (Du Plessis *et al.*, 2013:4). Since CAPS also mentions the idea that AL should be able to assist a learner to have access to further or higher education or the world of work, a third level of proficiency must be added, which is the economic level. Because this third level relates to the gaining of access to higher education and the world of work, it is also added to HL papers.

The CAPS outline is not restricted to language proficiency levels only, but also extends to differentiated language abilities which learners should master. This range of differentiated abilities stretches from educational and academic, aesthetic, political, economic, social and informational, as well as ethical contexts and situations.

In addition to the general aims of the curriculum explained above, CAPS (DBE, 2011a:9) identifies the following specific aims of language learning:

- language skills for academic learning;
- reading and viewing, writing and presenting, listening, and speaking;
- appropriate use of language for the audience, purpose, and context;
- expression and justification of learners’ ideas, views and emotions, orally and in writing;
- usage of language by learners in order to learn about themselves and the world around them;
- usage of language by learners in order to access knowledge and manage information for learning across the curriculum in a range of contexts; and
- usage of language by learners in order to think critically and creatively in a wider range of texts.

Section 3 of the CAPS outlines the content to be covered in language teaching. It is clear from this outline that the content is aimed at enabling learners to have a mastery of language in a wide range of different contexts and situations. As can be expected,

all these aims and content of the language also inform the assessment of the language.

In an attempt to describe the underlying construct of CAPS, Weideman (2009:40) conceptualises the sociolinguistic spheres outlined above as material spheres. Since these material spheres are used in the context of language, they then get referred to as material lingual spheres. The material lingual spheres give rise to a differentiated classification of language types. The differentiated language types are both factual (phonological, lexical, syntactic, semantic, etc.) and normative (aesthetic, legal, economic, social, ethical, etc.) in nature in that they originate from subjective human lingual capacity as well as from objective, factual language from social spheres.

Du Plessis *et al.* (2013:8) outline the advantage of employing a systematic theoretical framework such as Weideman's above that pertains to material spheres when teaching and assessing language as a system that can accommodate attributes peculiar to different languages, including indigenous African languages. Du Plessis *et al.*'s (2013) assertion above is important for this study in that it acknowledges Weideman's conception of the existence of material spheres in the present Grade 12 HL syllabus and it includes the applicability of the inclusion of attributes of assessment belonging to different languages, including indigenous languages. In addition, Du Plessis *et al.* (2013) identify the following as dominant material lingual spheres to the teaching and assessment of respective languages from CAPS (Du Plessis *et al.*, 2017:108):

- Social (including inter-personal communication and the handling of information);
- Educational (including academic and scientific language and advanced language ability);
- Aesthetic (including the appreciation of literature and art);
- Economic (including the world of work and commerce);
- Ethical (including the critical discernment of power relations in discourse); and
- Political (including an appreciation of the values embedded in language use).

According to Du Plessis *et al.* (2013), the above-mentioned material lingual spheres can also be referred to as discourse types. These discourse types inform fields in which languages operate. By so doing, they also inform a wide range of text types that could be used to learn a language. These discourse types operate in a differentiated way within reading, writing, speaking and listening abilities in Grades 10-12. The discourse types (or text types) can be integrated across different fields or spheres of discourse with regard to teaching. The most significant assertion made by these authors regarding the ability to be assessed by CAPS is that Grade 12 learners should have the ability to operate at different levels of proficiency in these diverse spheres (Du Plessis *et al.*, 2013:9-11). In addition, CAPS also covers what learners should be able to learn with regard to reading and viewing, as well as writing and presenting (DBE, 2011a:22-40). These abilities, according to Du Plessis *et al.* (2013:11), “should be considered as a full set of test specifications from which a selection has to be made when articulating the blueprint for the examination of language ability”. This is particularly relevant for the current study towards an identification of the construct upon which HL examination papers are based.

If reading and viewing, as well as writing and presenting, are identified as abilities necessary for the learners to gain competency in mastering the skill of accessing all the sociolinguistic spheres, then a deeper understanding of the operations of these abilities is necessary so as to ensure that important aspects of the construct for assessment are not omitted. According to Du Plessis *et al.* (2013:12), specific aspects in CAPS that involve reading and viewing are: the reading process, interpretation of visual texts, vocabulary development and language use, and sentence structures and the organisation of texts. The next section will explain each of these aspects in more detail.

4.4.1 The reading process applicable to all text types

The reading process that is applicable to all reading text types, both for HL and FAL, normally includes a three-phase process where pre-reading, reading and post-reading strategies are followed. The reading strategies alluded to here include strategies such

as scanning, skimming, visualising, predicting, evaluating, drawing conclusions, distinguishing between facts and opinion, and making meaning. Du Plessis *et al.* (2013:12), however, make a very critical observation about an anomaly that occurs in CAPS. They observe that CAPS still makes use of an outdated notion of separating the skills of language use (listening, speaking, reading and writing). While a separation of such skills makes sense in terms of their description, these skills should be integrated when one considers how they are used in reality. A good example of this is where summarising is recommended for assessment as a reading skill, while in actual fact it involves the integration of both reading and writing.

Another important observation regarding the reading process as contained in CAPS is its inclination to afford critical language ability considerable prominence. Critical language ability does not only address the issue of advanced language ability, but also the issue of other kinds of knowledge such as contemporary issues, human rights and environmental concerns. For this reason, the following list, adapted from a report by the original project team for the equalisation of Grade 12 examination papers, is included below as it outlines reading tasks requiring critical awareness and advanced language ability as espoused by CAPS (Du Plessis *et al.*, 2013:13).

Table 4.2: “Reading tasks requiring critical awareness and advanced language ability” (CAPS, DBE, 2011a:23) [adapted from Du Plessis *et al.*, 2013:13])

Critical language awareness involves the ability to:

- Distinguish between facts and opinions
- Understand direct and implied meaning
- Understand denotation and connotation
- Be aware of the socio-political and cultural background of texts and authors
- Notice the effect of selections and omissions on meaning
- Understand the relationship between language and power
- Recognise emotive and manipulative language, bias, prejudice, discrimination, stereotyping
- Recognise language varieties

- | |
|--|
| <ul style="list-style-type: none">• Recognise inferences, assumptions and arguments• Identify the purpose of including or excluding information |
|--|

One of the possible reasons why critical language ability is included in CAPS, could be to help provide access to higher education and encouraging an active and critical approach to learning as opposed to the much-criticised rote and uncritical learning. Since this ability is included in CAPS, a section on critical language ability should feature prominently in the Grade 12 examination papers. TALA tasks include this aspect (critical language ability) in the manner in which they are designed in the test. This ability could help to empower learners to display this skill and make first-year learners competent in critical thinking as opposed to the most lamented lack of critical skill at first-year level at university by learners fresh from secondary school.

4.4.2 Interpretation of graphic and visual texts

The abilities for the interpretation of graphic and visual texts sometimes go beyond just what learners need to learn at Grade 12-level. They even permeate communication science as is illustrated in Table 4.3 below taken from CAPS. As a result of the presence of economics as a lingual sphere in which learners need to achieve competency, the interpretation of graphic and visual texts is an important ability for learners to master. Aspects such as camera or font type as part of the present syllabus in Grade 12 are less relevant when compared to interpretation of graphic and visual texts – which are aspects of the TALA – and have nothing to do with language ability assessment, which are taught at Grade 12. Moreover, the academic literacy component (of which the TALA forms part as explained) as espoused by scholars such as Weideman (2017:119) includes 10 main subcomponents, and one of these main subcomponents includes “graphic and visual texts”.

Table 4.3: “Generic abilities for the reading and interpreting of visual texts” (CAPS, DBE, 2011a:23 [adapted from Du Plessis *et al.*, 2013:14]).

Learners should be able to: • 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 • Understand how language and images reflect and shape values and attitudes • Identify 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
--

4.4.3 Vocabulary development and language use

The assessment of the knowledge of vocabulary and language use is done in relation to text types used. It was mentioned earlier that text types used in Grade 12 learning are informed by sociolinguistic spheres. Since CAPS requires that learners should be able to explain the use of figurative language and rhetoric devices in different texts, these text types should be authentic examples of different languages. It was discovered, however, that most of the texts used for indigenous African languages in Grade 12 are not original but translated. In addition to this, most of the terminology used for indigenous languages is limited for these languages. In order to try and close the gap between these languages and English, substantial innovation to produce original indigenous African languages texts is necessary. Valid translation methods for translating texts is one of the approaches that could be used. Consequently, scientifically accepted methods of adaptation could be used for creating vocabulary and language items for tests, including the TALA in this study.

4.4.4 Sentence structure and the organisation of texts

CAPS lists “sentence structure and the organisation of texts” under reading and writing. This is, however, problematic due to the fact that when assessing reading, one has to also assess writing at the same time as separate abilities. There is, however, also strong encouragement from Du Plessis *et al.* (2013) that elements such as listening and speaking, and reading and writing, be integrated and not treated as separate skills. Du Plessis *et al.* (2013) point out that these skills are always intertwined in language use. Some of the examples necessary for the activity of language structure and and text organisation are listed in Table 4.4.

Table 4.4: “Examples of what learners should understand in terms of language structures and text organisation” (adapted from Du Plessis *et al.*, 2013:15)

Learners must be able to identify, explain and analyse the following language structures and conventions in texts:

- 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

4.4.5 Other components of assessment

There are other components of assessment that form part of Grade 12 examination papers that have not been discussed in the above outline of components that are included in such papers. Most of those components belong to school-based assessment (SBA) and writing and presenting. A discussion of such components are not included here because they do not form part of the 20% content envisaged for this study.

4.4.6 General underlying construct of HL examination papers

From the above discussion of the components of assessment, there is a strong notion that, according to CAPS, language assessment is aimed at addressing differentiated and generic abilities of language. It further creates the impression that language assessment is aimed at very heterogeneous types of language and that language is dynamic and constantly changing. CAPS captures this notion effectively but elaborately in its deliberations about what should be taught and assessed. However, the above discussion presented an outline of the most important aspects of CAPS and does not mean to be an exhaustive discussion of everything entailed by the document. After a close evaluation of how the underlying construct of a Grade 12 HL examination paper should be worded, the project team concluded that this construct is generic in nature and should be expounded as follows:

The assessment of 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 *et al.*, 2013; Du Plessis, 2017:118).

This underlying construct is important in that it projects effectively what should be the primary aim of the examination paper. Based on these deliberations, construct validity and equivalence between the different TALA tests employed in this study, i.e. the original English TALA and two Sesotho TALAs, must be ensured. According to Bachman and Palmer (1996:68), construct validity could be ensured by testers making explicit the nature of the ability that one wants to measure. This should include, among others, grammatical, textual, functional and sociolinguistic knowledge, strategic competence, and background knowledge. The most important question at this point, however, is whether the present Grade 12 examination papers are grounded in the common construct contained in CAPS across all 11 home language papers. Based on the definitions of construct validity by Cumming and Berwick (1996:1-12), Hughes (2003:26-35) and Weir (2005a:11-15), Du Plessis (2017:73) came to the conclusion that construct validity is a characteristic of both the cognitive and linguistic property,

arguably measured, and how it is aligned with various theories of how language is processed and communicated. As already mentioned in Chapter 1 of this study, based on the evaluation by several ministerial task teams as well as other observers, the African HL examination papers do not conform to the same construct in terms of the cognitive ability required compared to those of the English and Afrikaans HL examination papers. This leaves the African HL examination papers devoid of this specific feature in terms of sharing the common construct with English and Afrikaans, which brings into question the validity and reliability of the African HL examination papers.

4.5 Conclusion

Since the main objective of this study is to contribute towards the equalisation of Grade 12 African HL examination papers with the Afrikaans and English HL examination papers, literature about language testing was explored. The literature review included three important perspectives about the concept of validity, the validation process and a critique based on the project team's analysis of literature pertaining to construct validity as utilised by the South African NCS's CAPS document. Literature about different perspectives of validity that were identified by this study related to issues of the orthodox, modernist, and contemporary perspectives, as well as literature on validation explored issues of *a priori* and *a posteriori* validation. The discussion of the literature that focuses on CAPS explored the common underlying construct within language assessment in terms of CAPS and how that underlying construct validity should be articulated.

With relation to the evolution of the concept of validity, it was indicated that the concept of reliability encompasses a large part of the concept of validity. It also showed from various perspectives how validity evolved from its conceptualisation as a diverse disjunctive concept to its conceptualisation as a unitary concept, up to the point where, currently, only essential qualities of a test are emphasised for test validation. Therefore, prominent aspects identified in this literature survey were considered with regard to a comparison of the Translated Sesotho TALA with the original English TALA

in order to test its adequacy. As indicated in the chapter, these aspects have further been taken into account in the creation of the new Sesotho TALA, as well as in determining the adequacy of the Newly Created Sesotho TALA in comparison with the original English TALA.

CHAPTER 5: RESEARCH METHODOLOGY

5.1 Introduction

This study employs a quantitative research design, mainly because it deals with data in the form of numbers. In addition to this, it uses mathematical calculations to investigate the properties of data. It therefore makes use of quantitative analysis procedures, such as measurement, comparisons, examining relationships between tests, making forecasts, testing hypotheses, and explaining (William, 2011:113). This chapter illustrates how factors pertaining to the requirements of a quantitative research design were applied to this study.

It was necessary to apply the above-mentioned research design because the aim of this study, as mentioned in Chapter 1, is to create two different types of tests, one Translated Sesotho TALA (from the English version) and one Newly Created Sesotho TALA. In terms of this research design, these two tests measured against each other to determine which between the two is more adequate and appropriate for the purpose of this study, which is to create equivalence between HLs within the South African DBE as outlined in Chapters 2 and 4. This research also involves testing a hypothesis that if one of these tests would be found to be adequate and appropriate based on its internal and external validity, it would be used as a section comprising 20% of the 300 marks of the Grade 12 HL examination paper, to create a considerable level of equivalence between the Sesotho HL and English HL. Based on the outcome of the research whether the hypothesis is proven to be true or not, a recommendation was made. In line with one of the requirements for the quantitative research design, which is to elicit future predictions for the HLs, this study shall have achieved its objective.

This chapter firstly elaborates on how permission to conduct the empirical investigation at various schools, which formed part of this study, was obtained. School stakeholders from which permission was requested, included School Governing Bodies (SGB), School Management Teams (SMT), District Departments of Basic Education and the Provincial Department of Education. The procedure of acquiring permission included

information on how ethical clearance was obtained from the Ethics Committee for Language Matters (ECLM) at the NWU.

Lastly, this chapter explains how the empirical investigation was conducted. It incorporates information about the research questions that guided the investigation and how the empirical investigation was employed to address these questions adequately. The discussion therefore also includes a focus on the preparation of stimulus materials, selecting participants and administering the tests.

The main research question posed by this study focuses on the issue of test adequacy in terms of which test between the Translated Sesotho TALA and the Newly Created Sesotho TALA empirically shows the most significant internal and external validity. To address this question, the following sub-questions needed to be answered:

1. When considering both the Translated Sesotho TALA and the Newly Created Sesotho TALA , which of these two TALAs shows better evidence of internal validity with the English TALA?
2. When considering both the Translated Sesotho TALA and the Newly Created Sesotho TALA , which of these two TALAs shows better evidence of external validity with the English TALA?

In order to answer these sub-questions, relevant issues about the responsible design of valid and reliable tests were addressed, as well as what internal and external validity entail. With regard to internal validity, matters related to construct, face and content validity, as discussed in Chapter 4, were considered. External validity, on the other hand, focuses on predictive, concurrent and consequential validity that were also reviewed in Chapter 4. As indicated in Chapter 4, the important issues relating to internal and external validity that were highlighted in the chapter guide the design of valid and reliable tests. This is particularly important for both the design of the newly created TALA as well as the translation of the English TALA into Sesotho.

Furthermore, for the translated TALA, a theoretically sound approach for translating tests was incorporated. This is discussed comprehensively further on in the chapter.

5.2 Ethical considerations

Scholars agree that research should protect participants who volunteer to take part in it. According to Babbie and Mouton (2004:522), for example, research should not injure participants who take part in it. Similarly, Mackey and Gass (2005) insist that research should not harm participants. In order to ensure that the research for this study was not going to cause any harm to the participants, the researcher obtained ethical clearance from bodies that are entrusted with the authority to provide such clearance.

5.2.1 Acquiring permission from authoritative bodies

The researcher sought and obtained ethical clearance for the study from the North-West University's Ethics Committee for Language Matters (ECLM). The ethics number is NWU-00620-19-A7. After acquiring permission from the ECLM, further measures were taken.

Permission was further obtained from other bodies such as the Free State DBE. This was done in order to allow the researcher to write letters to the principals of schools, SGBs and SMTs to request permission for conducting research at their schools. Permission from the Department was granted and letters were written to all the principals, SGBs and SMTs of participating schools. The Free State DBE granted permission for all schools and the district (see appendix B2).

5.2.2 Selection of participants

Participants in this study were selected from four former Department of Education and Training (DET) secondary schools in the Sasolburg township of Zamdela in the north of the Free State Province. Learners from these schools were selected because all these schools have learners who belong to the cohort of learners who do Sesotho HL as a subject and their ages ranged between 17 and 18. Moreover, it was possible to select learners from all secondary schools in the same township. Selecting all learners from the same township ensured that a population belonging to this cohort was

possible. Selecting learners of ages ranging from 17 to 18 ensured that a model test similar to TALL test - which is a test meant for first-year students at university – could be designed easier. The population is made up of Grade 12 HL learners only. Ages for the participants in this study ranged from 17-18 years. These learners signed consent forms in which they agreed that they were going to take part in this study out of their own volition. They were, therefore, not coerced into taking part in the study. It was made clear in the consent form that participants were at liberty to withdraw from the study at any time if they wished to do so. Parents of participants in this study also signed assent forms in which they gave permission for their children to take part. Copies of both the consent and assent forms can be found in Appendix B3 of this study. The secondary schools from which the population derived are Nkgopoleng, Iketsetseng, Cedar and Sakhubusha. The language profile of learners from these secondary schools is that they are Sesotho HL and English FAL users. The number of learners doing Sesotho HL in each secondary school was determined before the practical implementation of the experiment (see Table 5.1).

Table 5.1: Numbers of learners who enrolled for the end of year examinations at participating schools

NAME OF THE SCHOOL	TOTAL NUMBER OF GRADE 12 LEARNERS
Nkgopoleng Secondary School	162
Cedar Secondary School	130
Iketsetseng Secondary School	170
Sakhubusha Secondary School	152
TOTAL	614

The following dates were set aside for the implementation of the experiment:

- Nkgopoleng Secondary School – 30 July and 5 August 2019;
- Cedar Secondary School – 31 July and 6 August 2019;
- Iketsetseng Secondary School – 1 August and 12 August 2019; and
- Sakhubusha Secondary School – 2 August and 8 August 2019.

As can be expected in a study like this, not all of the possible participants accepted the invitation to take part in the experiment. Attendance during the experiment varied according to different days and sessions as illustrated in Table 5.2. Certain learners, it would seem, took the liberty not to write some sessions. This is evidenced by varying numbers of attendance when scripts were counted. The outline provided in Table 5.2 is indicative of what actually transpired during different days and sessions at different schools.

Table 5.2: Numbers of learners who wrote the tests

NAME OF THE SCHOOL	DATE ON WHICH THE EXPERIMENT WAS SCHEDULED	TYPE OF TEST	NUMBER OF LEARNERS WHO WROTE THE TEST
Nkgopoleng Secondary School	30/07/2019	T-SES-TALA part 1	137
		T-SES-TALA part 2	134
		T-SES-TALA part 2 (4 questions that were answered on a separate score sheet)	92
	05/08/2019	NC-SES-TALA part 1	158
		NC-SES-TALA part 2	151
		NC-SES-TALA part 2 (4 questions that were answered on a separate score sheet)	92
Cedar Secondary School	31/07/2019	T-SES-TALA part 1	288
		T-SES-TALA part 2	120
		T-SES-TALA part 2 (4 questions that were answered on a separate score sheet)	121
	12/08/2018	NC-SES-TALA part 1	121
		NC-SES-TALA part 2	122
		NC-SES-TALA part 2 (4 questions that were answered on a separate score sheet)	119
Iketsetseng Secondary School	01/08/2019	T-SES-TALA part 1	128
		T-SES-TALA part 2	145
		T-SES-TALA part 2 (4 questions that were	116

		answered on a separate score sheet)	
		NC-SES-TALA part 1	123
		NC-SES-TALA part 2	119
		NC-SES-TALA part 2 (4 questions that were answered on a separate score sheet)	107
Sakhubusha Secondary School	02/08/2019	T-SES-TALA part 1	141
		T-SES-TALA part 2	143
		T-SES-TALA part 2 (4 questions that were answered on a separate score sheet)	123
	08/08/2019	NC-SES-TALA part 1	135
		NC-SES-TALA part 2	135
		NC-SES-TALA part 2 (4 questions that were answered on a separate score sheet)	119

Based on the above numbers, an average of 140 learners per school were able to write T-SES-TALA parts 1 and 2 on the stated dates and an average of 125 learners per school were able to write NC-SES-TALA parts 1 and 2 on the stated dates. The average number of learners who participated was reduced by an average of 21,5 per school from what was expected at the beginning of the study. The sample, however, remains significant in size.

Nkgopoleng Secondary School, Cedar Secondary School, Iketsetseng Secondary School and Sakhubusha Secondary School mentioned in the above statistics will be referred to as school 1, 2, 3, and 4 respectively in the remainder of the study.

5.3 Preparation of the stimulus material

The stimulus material in this study included both a translated test and an originally created test. Other material that was provided for this study besides the above-mentioned tests included answer sheets, erasers, and pencils.

Since the translation of tests has become a viable option in making tests available in more than one language, many adapt or translate tests as opposed to creating tests anew (Hambleton & Patsula, 2000:4). According to Hambleton and Patsula (2000:4), advantages of doing so include the low costs of adapting tests as opposed to creating new ones, “similarity of research findings between the original and adapted tests strengthens evidence for the validity of the test in its adapted form”, and cross-national, and/or cross-ethnic comparative studies require adapted tests. For this reason, this study wishes to determine whether a multilingual department of education such as the DBE would benefit most from translated tests or newly created tests for the purpose of potentially making use of the TALA in Grade 12 final examination papers.

Certain scholars add strong support for creating tests anew rather than translating or adapting tests. Some of the reasons which they advance, include the wrong perception by some test designers that adaptation will automatically result in valid tests. Hambleton and Patsula (1998:155) mention that it is not always true that adaptation results in valid tests – they also posit that developing a new test may increase the validity of a test over the level of validity obtained by an adapted test. Another important reason why certain researchers opt for newly created tests, is the risk of researchers imposing conclusions on concepts based on their cultures which are foreign when viewed from the point of view of other cultures (Hambleton & Patsula, 1998:156).

Based on the two opposing views discussed above, the aim of this study is, therefore, to determine which of the two tests, i.e. the Translated Sesotho TALA and the Newly Created Sesotho TALA, that were developed for this study, performed best when scientifically measured against each other according to language testing principles.

Since the TALAs are language tests meant to be used for assessment of proficiency and achievement of advanced language ability for school learners, it was deemed necessary to evaluate if they could be used as part of Grade 12 HL examination papers. As mentioned previously, this study aims to extend the potential use of the TALA to African HL examination papers (with specific reference to Sesotho). At present, these tests exist as English and Afrikaans HL tests developed by other researchers for a purpose similar to this study (Steyn, 2018). The following two sections of the chapter deal with the specific approach that was followed for the translation of the English TALA into Sesotho and the design and development of the Newly Created Sesotho TALA.

5.3.1 Translation of the English TALA into Sesotho

As explained before, the translation of the English TALA into Sesotho was just one of the two strategies used for this study, the other being the development of a Newly Created Sesotho TALA. The aim was therefore to determine which of the two strategies would produce the highest level of equivalence with the English version of the test. Thus, this investigation had to inform our understanding of which of the two strategies was the most appropriate and adequate for replacing a portion (20%) of the Grade 12 HL examination paper. The translation of the English TALA into Sesotho was therefore approached from the outset to adequately address the issue of equivalence between the English TALA and the Translated Sesotho TALA. The first important matter was that the Translated Sesotho TALA had to stay true to the construct on which the English TALA was based. It further had to meet all the requirements for a valid and reliable language test as explained below under the ITC guideline, especially the second pre-condition guideline (“the amount of overlap in the definition and content of the construct measured by the test and the item content in the populations of interest should be evaluated in order to determine that it is sufficient for the intended use of the score” (ITC, 2016:9). In addition to this, this translation was supposed to happen according to various requirements for a successful translation as espoused by recognised translation bodies, such as the guidelines of the International Test Commission (ITC, 2016), so as to contribute to the validity of the test. How the

109

validity of the test can be achieved by translation will be explained below under the guidelines intended for that purpose. Moreover, as already mentioned with regard to the model of translation, the functionalist approach was used.

As mentioned in Chapter 3, the functionalist approach would be applied extensively in the translation of the English TALA into Sesotho (these two TALAs from now on will be labelled ENG-TALA part 1 and part 2 [ENG-TALA part 1 & 2] and Translated SES-TALA part 1 and 2 [T-SES-TALA part 1 & 2], respectively). Parts 1 and 2 of the TALAs contain different sections of the same test. Part 1 of each of the TALAs refers to the first part of the TALA that tests scrambled text and text comprehension and is made up of 83 marks in total. Part 2 of each of the TALAs includes the sections that tests graphic and visual information, vocabulary knowledge, and grammar and text relations and is made up of 104 marks in total. A complete analysis of what each of these parts contains will be presented below. It was also mentioned in Chapter 3 that the functionalist approach to translation is based on three functions from Karl Buhler's organon model (Nord, 1997:40), including one additional function from Nord (1997). Examples of how each of the functions was applied in the translation of the ENG-TALA part 1 and 2 into the Translated SES-TALA part 1 and 2 were also provided in Section 3.3 – "Nord's functionalist approach to translation". This included the procedure of adaptation as well as other strategies such as direct and literal translation. Most of the requirements for a valid translated test as hereby expounded were applied to the template of the original English TALA which was used for the Translated SES-TALA. The common construct thereof is the one agreed upon by the project team. This was already presented in Chapter 1 as:

The assessment of 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 *et al.*, 2016:118).

The template that was generated based on this construct includes the following major elements of the TALA as contained in the above common construct:

- the language ability to be tested;

- discourse types;
- texts; and
- functional and formal aspects of language.

The ENG-TALA part 1 contains the following sections:

- Section 1: scrambled texts (“For the love of strawberries” and “MSG – food baddies or misunderstood enhancer?”) (10 questions), and
- Section 2: comprehension questions testing vocabulary, understanding metaphor and idiom, textuality (cohesion and grammar), understanding text types (genre), distinguishing between essential/non-essential information, and making meaning beyond the sentence (73 questions).

The same sections mentioned above for the ENG-TALA part 1 were, for obvious reasons, utilised for the T-SES-TALA part 1 with the same number of questions.

The ENG-TALA part 2 contains the following sections:

- Section 1: interpreting graphs and visual information (titled “World grain trade and production”) testing vocabulary, distinguishing between essential/ non-essential information, making meaning beyond the sentence, numerical computation, and extrapolation, application and finding evidence for an argument (30 questions);
- Section 2: vocabulary questions (32 questions); and
- Section 3: cloze-type questions testing vocabulary and textuality (cohesion and grammar) (42 questions).

Again, the same sections mentioned above for ENG-TALA part 2 were utilised for the T-SES-TALA part 2 with the same number of questions. In order to visually illustrate the different sections of both the ENG-TALA and the T-SES-TALA part 1 and part 2 as explained above, Table 5.3 was generated. Table 5.3 illustrates sections belonging to ENG TALA or T-SES-TALA part 1, and ENG-TALA or T-SES-TALA part 2.

Table 5.3: Divisions of tasks for different question types

ENG-TALA/T-SES-TALA part 1	Tasks	Number of Questions
Section 1	Scrambled text: “For the love of strawberries” // Bakeng sa lerato la distrawberry//	10
Section 2	Comprehension:	73
	TOTAL	83
ENG-TALA/T-SES-TALA part 2	Tasks	Number of questions
Section 1	Interpreting graphs and visual information (titled “World grain trade and production” // Kgwebisano ya Lefatshe ya dijothollo le tlhahiso//	30
Section 2	Vocabulary questions	32
Section 3	Cloze-type questions	42
	TOTAL	104

With regard to the translation/adaptation of the ENG-TALA part 1 and 2 into Sesotho, the ITC guidelines served as a functional instrument towards the achievement of that objective. Examples of how these guidelines were applied are included below.

It is important to note that the ITC document states that these guidelines are meant for the adaptation of psychological and educational tests for use in different populations (ITC, 2016:8). However, the application of the ITC guidelines is not limited to psychological and educational tests only, they are also used for language tests (ITC, 2016:11).

According to the ITC document (2016), the 18 guidelines are divided as follows:

- Pre-Condition (PC) (3 in number);
- Test development (TD) (5 in number);
- Confirmation (C) (4 in number);
- Administration (A) (2 in number);
- Score scales and interpretation (SSI) (2 in number); and
- the document (D) (2 in number).

These guidelines are updated periodically by well-qualified experts and have been derived from a wide range of research as can be seen from periodical ITC reports. The next section will deal with the three pre-condition guidelines first.

5.3.1.1 Pre-condition guidelines

The first guideline under the pre-condition guidelines is PC-1. It states that “permission from the holder of the intellectual property rights to the test should be obtained before carrying out any adaptation” (ITC, 2016:8). For this reason, permission from the researcher, Sanet Steyn, whose test was to be translated was sought and permission was granted.

The second guideline under the pre-condition guidelines is labelled PC-2. This guideline states that “the amount of overlap in the definition and content of the construct measured by the test and the item content in the populations of interest should be evaluated in order to determine that it is sufficient for the intended use of the score” (ITC, 2016:9). This implies that experts who are also familiar with the cultural groups that are being tested should evaluate the legitimacy of the construct measured in each linguistic group.

For the above-mentioned requirement, two individuals, Lebogang Mathibela and Lebona Mafisa, who were familiar with the construct measured in each cultural group and who had undergone two sessions of training in the construct, were recruited. These experts hold language practice (translation and interpretation) and education

qualifications. Their training include language testing aspects by qualified applied linguistic experts. They were able to determine if the construct of the Sesotho version of the test was similar to the English version. They were also able to resolve whether the construct of the two tests would make sense for the two linguistic groups.

The third and the last guideline (labelled PC-3) under the pre-condition guidelines states that “the influence of any cultural and linguistic differences that are irrelevant to the intended uses of the test in the population of interest should be minimised” (ITC, 2016:10). This implies that the cultural and linguistic characteristics irrelevant to the variables that the test is intended to measure should be identified at an early stage of the project (ITC, 2016:10).

Based on the above explanation, the two individual experts were tasked to identify such elements that they thought could apply and to note them so as to minimise bias. They observed that texts from the original test (i.e., ENG-TALA part 1 and part 2) would culturally be Western-based and foreign to most of the Sesotho learners, especially at Grade 12 level. For instance, titles for texts are meant to be catchy according to the language in which they originate, but could be completely foreign in a target language if literal translation is applied. To overcome this obstacle, either an equivalent expression or meaningful translation needs to be effected. As a means of illustration, ENG-TALA part 1 will be used.

The title of the first text used in the ENG-TALA part 1 is “For the love of strawberries”. This title was translated as “*Bakeng sa Lerato la distrawberry*” in the T-SES-TALA part 1. This was neither a literal nor an equivalent translation of the title. Some critics could argue that an equivalent word for strawberry does exist in Sesotho, but a word that people normally refer to as such is general and not specific. The word *monokotshwai* that is used in Sesotho refers to different types of berries and not ‘strawberry’ specifically. In order to overcome this confusion, the word strawberry was foreignised to preserve its meaning. This strategy was applied extensively in the translation of the ENG-TALA part 1 and 2 to SES-TALA part 1 and 2 in order to effect the same objective as the one illustrated here. There are, however, other technicalities at play for this

strategy which had to be addressed. Because of the tendency of Sesotho to sometimes domesticate certain words, some translators might expect the word 'strawberry' to be written according to Sesotho orthography, something like 'stroberri'. However, this is also problematic if it is the first time an author writes this, as sometimes the manner in which these words are written is different. Long-time use entrenches the best way of writing a foreign word like that.

5.3.1.2 Test development guidelines

The first sub-guideline labelled TD-1 under the test development guidelines states that "the translation and adaptation processes of linguistic, psychological, and cultural differences in the intended populations through the choice of experts with relevant expertise should be ensured" (ITC, 2016:11). With the expertise already defined for the choice of experts for this project, it is clear that experts used in this project met the required criteria. These two experts are not just translators, but they are both educators and people who know the culture of the target group as required by the guidelines.

The second sub-guideline (TD-2) under test development states that "appropriate judgemental designs and procedures to maximise suitability of the test adaptation in the intended population should be used" (ITC, 2016:12). The most popular judgemental designs, according to the ITC (2016:12) specifications, are forward and backward translation designs. Forward translation implies translating a text into a TL, while backward translation implies translating the translated text backward through back translation. According to Humbleton and Patsula (2000:14), back translation for judgemental design is popular, but forward translation designs provide stronger evidence for test equivalence because both the SL and TL versions of the test are scrutinised. In addition, they assert that a test that can be back translated correctly (backward), is not a guarantee of the validity of the TL version of the test. The reason for this is that if the backward translation design is implemented in its narrowest form, no review of the TL version of the test is ever done. Sometimes this design leads to an acceptable TL translation that makes it easy for back translation, but sometimes it

leads to an awkward TL translation version of the test. The proper way of implementing this design is through making sure that back translation is done and changes are made to the adapted test, and that the forward translation is implemented. For the translation of both the ENG-TALA part 1 and 2, two translators were commissioned to forward translate the test and two different translators were commissioned for its back-translation. The researcher, who is a language practitioner himself, and the two translators who were commissioned with the duty of forward translation, convened to make changes to the adapted test. Judgemental reviews were done independently by each of the translators and decisions were taken unilaterally by the researcher and the two translators about the final product.

The third guideline under test development guidelines (TD-3) states that “evidence that the test instructions and item content have similar meaning for all populations should be provided” (ITC, 2016:14). Since this guideline requires that translators have some knowledge of test construction and the testing culture in the target group, as explained in the section above on “How the guideline of the International Test Commission (ITC) was applied”, the two translators who formed the committee for the translation of both ENG-TALA part 1 and 2 were used for this purpose. The researcher was, therefore, able to use the judgemental methods necessary for adaptation such as the use of reviewers with expertise in local culture and language, use of samples of bilingual respondents, use of local surveys to evaluate the test, and use of non-standard test administrations to increase acceptability and validity. This was accomplished through the use of Sesotho educators at schools where the tests were administered to provide an overview assessment of the original English test and the translated Sesotho one. The response that was received from school teachers confirmed that the two parts of the translated test have a similar meaning with regard to both instructions and item content to the two parts of the original English test.

The fourth sub-guideline (TD-4) under the test development guidelines states that “evidence that the item formats, rating scales, scoring categories, test conventions, modes of administration, and other procedures are suitable for all intended populations should be provided” (ITC, 2016:15). Broadly speaking, ENG-TALA part 1 as well as

the translated SES-TALA part 1 are divided into two sections, namely, scrambled text and comprehension. There are two scrambled texts titled: “For the love of strawberries” and “MSG – food baddies or misunderstood enhancer?”. This is a task where sentences from a paragraph are scrambled, and a learner needs to arrange them according to a logical sequence. There are five sentences labelled A, B, C, D, and E in a paragraph. The requirement for the question is to look at five given numbers, each with a scale of five letters (e.g., 1. A, B, C, D, E) of the alphabet representing listed sentences and to choose one letter of the alphabet for each sentence number. The following screenshot (Figure 5.1) from ENG-TALA part 1 serves to visually illustrate this section.

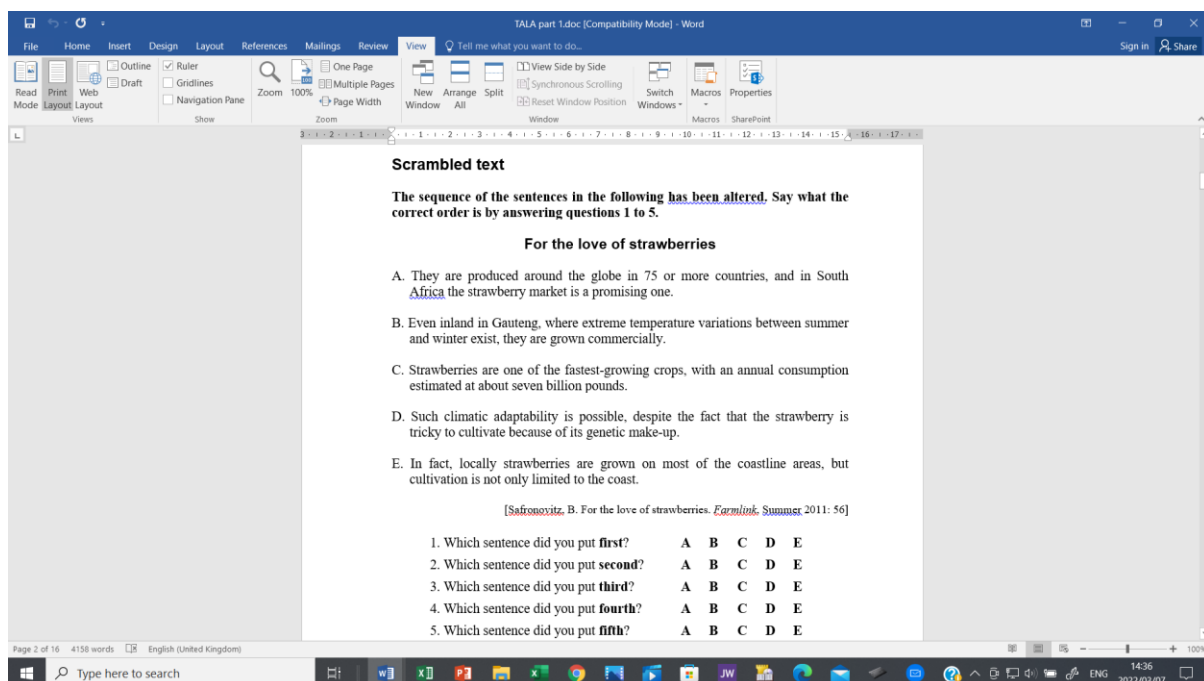


Figure 5.1: Screenshot of scrambled text question, ENG-TALA part 1

The same procedure is required for the second paragraph on scrambled text. In both the ENG-TALA part 1 and Translated SES-TALA part 1, the same procedure for questioning was used.

The comprehension part contains mixed questions derived from academic vocabulary, text comprehension, verbal reasoning, and register and text type. Each numbered question is a multiple-choice type question with four different options to choose from. The scale for answers ranges from A–D as can be seen in the example of a screenshot from the ENG-TALA (Figure 5.2)

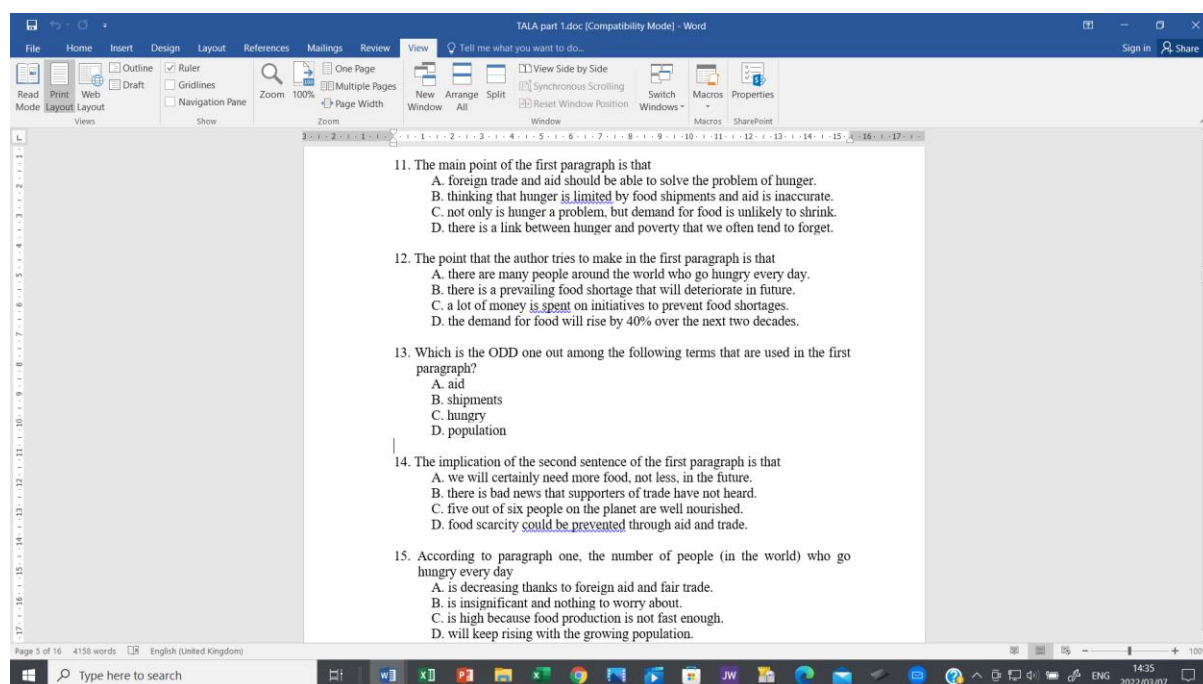


Figure 5.2: Screenshot of multiple-choice question with scale ranging from A–D for ENG-TALA comprehension questions

The test was written physically as a pen-and-paper exercise and not done electronically. Candidates had 55 minutes to complete the test. The 55 minutes time limit for this test was based on the time allocation of a similar test (the TALL) upon which this test is based. This applies for both the original ENG-TALA part 1 and the Translated SES-TALA part 1. The researcher took all possible care to ensure that item formats, rating scales, scoring categories, test conventions, modes of administration, and other procedures were suitable for all intended populations as demonstrated in this paragraph.

The fifth and the last sub-guideline under test development guidelines (TD-5) requires that “pilot data on the adapted test be collected in order to enable item analysis, reliability assessment and small-scale validity studies so that any necessary revisions to the adapted test can be made” (ITC, 2016:16). For this study, data was collected from all the schools at which the experiment was performed. This data was statistically analysed using *TiaPlus* and *Winsteps*. The results of this data and the revisions that were made based on the analysis will be discussed in more detail in Chapter 6.

5.3.1.3 Confirmation guidelines

The next type of ITC guideline is “confirmation guidelines” (ITC, 2016:17). This guideline is made up of four sub-guidelines (C-1 to C-4) necessary for empirical analyses rather than judgement. As the definition thereof explains, these sub-guidelines will not be discussed in this chapter. These guidelines will, therefore, be discussed in Chapter 6 of this study which focuses on the analysis of data.

5.3.1.4 Administration guidelines

The next ITC guideline is termed “administration guidelines”. This is made up of two sub-guidelines, namely A-1 and A-2. A-1 deals with “the issue of preparing administration materials and instructions to minimise any culture and language related problems that are caused by administration procedures and response modes that can affect the validity of the inferences drawn from the scores” (ITC, 2016:25). As instructions about how to answer questions and which questions to answer first can be problematic for Grade 12 learners in the South African school setting, this issue was presented to learners in the simplest terms possible. The problem for this situation is caused by anxiety and lack of comprehension by learners in examinations. Test instructions were provided orally before the start of the test in order to allow learners time to ask questions which they might have. This was done step by step with learners who already, at this time, had access to the question papers and answer sheets. This part took just over ten minutes to complete. It was followed by another 20 minutes of instructions on how to complete the answer sheet. As both the translated SES-TALA

part 1 and part 2 were meant to be written on the same day, a period of one hour instruction was set aside for this purpose. All materials necessary for the administration of each of these TALAs such as pencils, erasers, question papers and answer sheets were supplied by the researcher and information about how to use them was thoroughly discussed with learners before they wrote the test.

The A-2 sub-guideline is meant “to specify testing conditions that should be followed in all populations of interest” (ITC, 2016:26). These testing conditions included conditions that each learner would spend 55 minutes to complete the test, hard copies would be exactly the same, and the test would be administered under more or less the same weather condition where there are no bad weather conditions. This was achieved for the Translated SES-TALA part 1 and 2 by requesting school principals to offer their examination centres as venues for the administration of these tests, just as the ENG-TALA part 1 and 2 were written in examination venues. Test instructions were the same for both the translated SES-TALA part 1 and 2 and the ENG-TALA part 1 and 2. A list of learners with special needs was requested, but it was discovered that no learner had special needs from the four different schools that were used for these tests.

5.3.1.5 Score scales and interpretation guidelines

The next test guideline is termed ‘score scales and interpretation guidelines’ (SSI). There are two sub-guidelines that form part of this guideline, namely, “the specifications of testing conditions that should be followed closely in all populations of interest (SSI-1) and the interpretations of any group score differences with reference to all available information (SSI-2)” (ITC, 2016:26-28). This guideline also focuses on test score interpretations and will, therefore, also be reserved for discussion in the data analysis in Chapter 6.

5.3.1.6 Documentation guidelines

Following the above-mentioned guidelines based on score interpretation, is the guideline termed ‘documentation guidelines’ (DC). This is made up of guidelines on

“providing technical documentation of any changes, including an account of evidence obtained to support equivalence, when a test is adapted for use in another population (DC-1) and documentation for users that will support good practice in the use of an adapted test with people in the context of the new population” (DC-2) (ITC, 2016:28-29). All these records are supplied in the discussion of the results in the analysis of the data (Chapter 6).

The 22 guidelines mentioned above are required by the ITC, as already mentioned. The successful translation of the test applied in this study does, however, not end with the implementation of the guidelines of the ITC only. It goes on to include an awareness of other principles of successful translation that were already discussed in Chapter 3 of this study. Explanations of how some of these principles work in this study were cited with examples when different key concepts of translation were reviewed in Chapter 3 (Sections 3.3 to 3.8).

It was further mentioned in Chapter 3 that translation procedures and strategies derived from different sources would be utilised in this study. In order to ensure a valid translation approach, information about procedures and strategies such as borrowing, calque, transposition, modulation, equivalence, amplification, explicitation and generalisation, and how these were applied in this study was supplied in the discussion of Vinay and Darbelnet’s model in Chapter 3. Back translations as important strategy was also elaborated upon in the same chapter and its application in this study was practically illustrated in Section 5.3.1.2.

In line with the requirements for a quantitative research design, the T-SES-TALAs were marked electronically by means of a scanner. This was easily accomplished as this test is a multiple-choice type of test. An answer to each question is denoted by a letter of the alphabet, i.e., A, B, C, D, and E. The results for each test were transferred to an *Excel* spread sheet according to the format reflected in Table 5.4. Reflected in this table (5.4), is an example of question number, labelled item number and correct answer according to the memorandum of questions. This example also shows student numbers and how each student answered each question or item.

Table 5.4: Example of items for T-SES-TALA part 1 & 2

Weight	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Memo	C	A	E	B	D	D	E	C	B	A	C	D	B	A	D	C	A
Task	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Student No.																	
1	C	B	D	E	A	D	E	A	C	B	B	X	B	E	D	A	D
2	C	E	A	B	D	D	C	E	A	B	A	B	A	B	D	D	B
3	C	B	E	D	A	D	C	B	A	E	C	B	C	A	C	A	A
4	D	C	E	B	A	D	E	A	C	B	C	D	C	A	C	A	A
5	C	A	E	B	D	D	C	A	E	D	A	B	A	C	D	B	A

The complete *Excel* document with all answers recorded was sent electronically to the researcher after scanning was completed by means of the scanner. These results were saved in a folder on the computer of the researcher.

This *Excel* spread sheet contains all the items of both the 83 items for part 1 and 104 items for part 2 of the test, bringing the total to 187 items. However, the answers or items were converted to numbers as illustrated below in Table 5.5. This data set therefore included all students who wrote this test, and is considered the second set of data for this study. This data set was sent to a statistician for analysis, using the *TiaPlus* software for the analysis in terms of item facility, item discrimination, overall reliability and subtest intercorrelations in order to measure its adequacy and appropriateness. The items in the *Excel* spread sheet were converted to numbers so as to enable *TiaPlus* to process them.

Table 5.5: Example of items for T-SES-TALA part 1 & 2

Weight	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Memo	3	1	2	2	4	4	5	3	2	1	3	4	2	1	4	3	1
Task	1	2	3	4	5	6	7	8	9	10	11	12	13	14	1	16	17
Student No.																	
1	3	2	4	5	1	4	5	1	3	2	3	3	2	5	4	1	4
2	3	5	1	2	4	4	3	5	1	2	1	2	1	2	4	4	2
3	3	2	5	4	1	4	3	2	1	5	3	2	3	1	3	1	1
4	4	3	5	2	1	4	5	1	3	2	3	4	3	1	3	1	1
5	3	1	5	2	4	4	3	1	5	4	1	2	1	3	4	2	1

5.3.2 Designing the original Sesotho TALA test

It is important in this section to keep in mind that the TALA is modelled on the TALL. In the development of the Sesotho TALA, the TALL as well as the Afrikaans version of the TALA called the *Toets van Gevorderde Taalvaardigheid* (TOGTAV) and the English TALAs were used as templates for design. Therefore, an exposition of how all three above mentioned tests of academic literacy were designed and how they influenced the design of the TALA will be explored in this section. Most of the information pertaining to the design of the TALL also applies to the development of the Afrikaans and English TALAs and the information pertaining to the design of the Sesotho TALA consequently derives from these three tests. Similar to the ENG-TALA part 1 and 2, the Newly Created Sesotho TALA was divided into two parts, labelled Newly Created SES-TALA part 1 and 2 (NC-SES-TALA part 1 & 2).

The NC-SES-TALA part 1 and 2 used the same blueprint as that which was used for the ENG-TALA part 1 and 2. A strong commitment not to deviate from the common construct as contained in the ENG-TALA was adhered to in the design of the NC-SES-TALA part 1 and 2. All major elements such as language ability to be tested, discourse types, texts and functional and formal aspects of language from the original template were retained. Thus, the NC-SES-TALA part 1 was divided into two scrambled texts titled “*Noka ya Senqu*” (The Orange River) and “*Mouwane wa metsi o ka fana ka matla a mangata a ntlafaditsweng*” (Steam can supply clean additional energy) which, when added together, made up 10 marks; and 73 comprehension questions that test vocabulary comprehension, understanding metaphor and idiom, textuality (cohesion and grammar), understanding text types (genre), distinguishing between essential/non-essential information, and making meaning beyond the level of the sentence. All 83 questions were based on the text titled “*Na metsi a rona a tswa hokae?*” (Where does our water come from?). Marks for the NC-SES-TALA part 1 added up to a total of 83 marks. No readability test exists for the Sesotho language.

However, the following parameters were generated when the same text was translated into English: Flesch Reading Ease – 54,8, Flesch-Kincaid Grade level 10,1. This readability parameter is appropriate for learners at Grade 12 in South African schools (Steyn, 2018:38).

The NC-SES-TALA part 2 contained the following sections:

Interpreting graphs and visual information

This is where the text titled “*Bokgoni ba ho itlhokomela ka ‘matla’ ka diperesente (%)*” (Ability to supply energy in percentages) was utilised for 30 questions. These questions tested vocabulary information, distinguishing between essential/non-essential information, making meaning beyond the sentence, numerical computation, extrapolation, application, and finding evidence for an argument.

Vocabulary knowledge

This part is based on the text titled “*Eskom le bothata bo tebileng ba metsi Afrika Borwa*” (Eskom and the serious water problem in South Africa). It is made up of 32 questions that test vocabulary and textuality, i.e. cohesion and grammar.

Grammar and text relations

This part is also based on the text titled “*Eskom le Bothata bo tebileng ba metsi Afrika Borwa*” (Eskom and the serious water problem in South Africa) and contains 42 cloze-type questions.

These two parts of the NC-SES-TALA are based on the theoretical foundations of TALL and made up a total of 104 marks. The design and the development of the TALL is based on three principles of design, according to Weideman (2011). These principles are, first, the theoretical defensibility of the definition of the ability being measured, or the construct on which the test is based; secondly, the consistency and stability of the test as a measuring instrument; and thirdly, whether and how one might come to valid and reliable interpretations of the test results.

As far as the first principle is concerned, “the theoretically defensible construct for the TALL had to be the one that conceived academic literacy as the ability to handle academic discourse at university” (Weideman, 2011:103). For TALA, the desirable and theoretically defensible construct would be the one that conceived of the TALA as the ability to handle the differentiated language ability in a number of discourse types involving typically different texts, and the generic ability incorporating task-based functional and formal aspects of language. The student would need to handle these abilities at Grade 12 level. This is in line with the principle of high knowledge and high skills as encapsulated in the CAPS document (Du Plessis *et al.*, 2013). This first principle of test design is in line with Bachman and Palmer’s assertion that tests should be based on an interactional perspective of language ability which emphasises the ability to negotiate meaning within specific contexts (Bachman & Palmer, 1996:61). As this first principle deals with the part of the construct focusing on “the ability to be measured”, this concurs with the definition of what a construct is as defined by Bachman and Palmer (1996). This leads to the designers of TALA arriving at the conclusion that this test, like the TALL, should measure learners’ ability in being competent with the negotiation of meaning in specific contexts. Learners should therefore display their competence with regard to the following abilities derived from Weideman (2011:103):

1. understand a range of academic vocabulary in context;
2. interpret and use metaphor and idiom, and perceive connotation, word play and ambiguity;
3. understand relations between different parts of a text, be aware of the logical development of (an academic) text, via introductions to conclusions, and know how to use language that serves to make the different parts of a text hang together;
4. interpret different kinds of text type (genre), and show sensitivity for the meaning that they convey, and the audience that they are aimed at;
5. interpret, use and produce information presented in graphic or visual format;
6. make distinctions between essential and non-essential information, fact and opinion, propositions and arguments; distinguish between cause and effect, classify, categorise and handle data that makes comparison; and
7. make meaning beyond the level of sentence.

These abilities pertaining to the TALL compare favourably with abilities that the CAPS document outlines as abilities necessary for addressing the principle of high knowledge and high skill. They compare favourably in that, although somehow differently stated, they address the same abilities. The CAPS abilities regarding the competency which learners should be able to display are outlined in Chapter 4 under the section “Theoretically analysing the usage of an underlying common construct for TALA”.

Table 5.6 outlines important abilities and related task types pertaining to the above-mentioned abilities contained in the TALL and the TALA. These abilities and tasks are commonly shared by both the TALL and the TALA for reasons already discussed.

Table 5.6: “Test components and/or specifications of a test of academic literacy” (Mayburg-Smit, 2017:278)

Specification/ component and / or of a test of academic literacy	Possible task types
Vocabulary comprehension	Vocabulary knowledge Dictionary definitions Cloze C-procedure
Understanding metaphor and idiom	Longer reading passages
Textuality (cohesion and grammar)	Scrambled text Cloze C-procedure (perhaps) Register and text type Longer reading passages Academic writing tasks

Specification/ component and / or of a test of academic literacy	Possible task types
Understanding text type (genre)	Register and text type Interpreting and understanding visual & graphic information Scrambled text Cloze procedure Longer reading passages Academic writing tasks (possibly also) C-procedure
Understanding visual & graphic information	Interpreting and understanding visual & graphic information (potentially) Longer reading passages
Distinguishing between essential/non-essential information	Longer reading passages Interpreting and understanding visual & graphic information Academic writing tasks
Numerical computation	Interpreting and understanding visual and graphic information Longer reading passages
Extrapolation and application; finding evidence for an argument	Longer reading passages Academic writing tasks (Interpreting and understanding visual & graphic information)
Communicative function	Longer reading passages (possibly also) Cloze, scrambled text
Making meaning beyond the sentence	Longer reading passages Register and text type

Specification/ component and / or of a test of academic literacy	Possible task types
	Scrambled text Interpreting and understanding visual & graphic information

In addition to those aspects that are common between the TALL and TALA presented above, another similarity exists between these aspects and the aspects included in the framework for language abilities proposed by Bachman and Palmer (1996). This framework is illustrated in Figure 5.4 derived from Van Dyk and Weideman (2004:8).

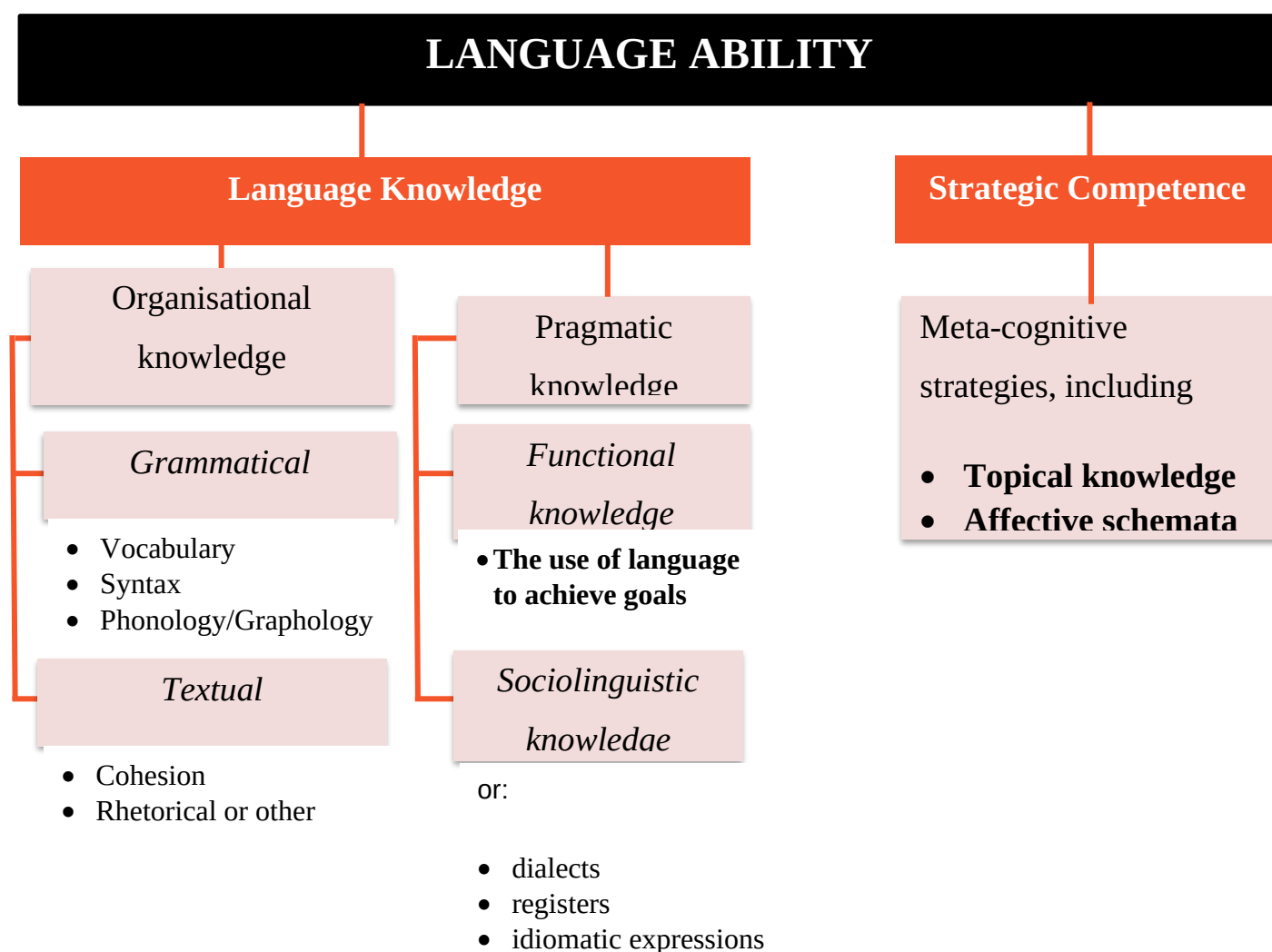


Figure 5.3: “The Bachman and Palmer language ability framework” (Van Dyk & Weideman, 2004a:8)

The following language abilities related to the above framework have been identified from the English TALA template.

1. Language knowledge:

- (a)** Organisational knowledge (knowledge of rhetorical or conversational organisation): e.g. scrambled text and comprehension questions (these appear prominently in the English TALA part 1, the Translated Sesotho TALA part 1 and the Original Sesotho TALA part 1).

- (b) Grammatical knowledge (vocabulary, syntax, and phonology): e.g. vocabulary, grammar, graphology and text relations (these with the absence of phonology appear prominently in the English TALA part 2, the Translated Sesotho TALA part 2 and the Original Sesotho TALA part 2).
- (c) Textual knowledge (cohesion, rhetorical or other): e.g. text relations (these appear in the English TALA part 2, the Translated Sesotho TALA part 2 and the Original Sesotho TALA part 2).
- (d) Pragmatic Knowledge: e.g. Knowledge needed to communicate effectively. As Du Plessis (2012:47) rightfully states, this knowledge also impacts upon knowledge of rhetorical or conversational organisation, which resorts under the category of organisational knowledge, and it is also related to communicative goals, as well as functional and sociolinguistic knowledge. This was already discussed in terms of (a) above where organisational knowledge is contained in the TALA. Communicative knowledge embraces different aspects in both language knowledge and strategic competence which are spread all over the different sections of the TALAs.
 - * Functional knowledge refers to the use of language to achieve a goal. This knowledge cannot easily be separated from textual knowledge and can therefore be categorised in the TALA as illustrated in (c) above.
 - * Sociolinguistic knowledge includes aspects such as dialects, registers, and idiomatic expressions. But like functional knowledge, it can also be categorised with textual knowledge (Van Dyk & Weideman, 2004:47) and the use thereof in TALA is illustrated in (c) above.

2. **Strategic competence** includes aspects such as meta-cognitive strategies like topical knowledge and affective schemata. Meta-cognition is defined as awareness about one's thought processes and in TALA (all the different versions used in this study), it is embedded in the selection of topical texts such as the ones that are used for these tests.

It was illustrated above regarding the principle of high knowledge and high skill for Grade 12 HLs, that there is a strong inter-relatedness between Weideman's

130

(2011:103) summary of abilities for academic literacy, Mayburg-Smit's TALL aspects and tasks (Mayburg-Smit, 2017:278), and language abilities outlined by Bachman and Palmer (1996) as illustrated by Van Dyk and Weideman (2004:8). The manner in which each one relates to the design of TALA was explored and practically illustrated. It is against this background that NC-SES-TALA was designed and developed.

Again, in line with the requirements for quantitative research design, NC-SES-TALAs have been marked electronically by means of a scanner and the Table 5.7 captures the data set for the NC-SES-TALA in alphabetical letters.

Table 5.7: Screen shot of items for the NC-SES-TALA part 1 & 2 in alphabetical letters

Memo	D	E	A	C	B	D	C	C	D	D	C	B	C	C	C	C	D
Item number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Student No.																	
1	D	A	C	B	B	C	B	A	C	D	A	C	A	A	B	C	A
2	B	A	B	A	B	B	C	A	C	D	B	B	A	C	C	B	A
3	D	A	B	C	E	D	C	C	D	D	C	B	C	B	A	C	A
4	E	A	B	C	D	C	C	C	D	D	C	B	C	C	C	C	D
5	B	E	D	A	C	C	C	C	D	D	C	B	B	C	C	C	D

Similar to the T-SES-TALA, the *Excel* script (Table 5.7) was produced after having been marked by the scanner and, similar to the data set for the T-SES-TALA, the information in the spread sheet had to be converted to numbers, as can be seen in Table 5.8.

The data set in Table 5.8 reflects the set that was sent to the statistician for analysis by means of the *TiaPlus* and *Winsteps* software packages. This data set will be compared with the data set for the T-SES-TALA in the analysis of the data in order to arrive at an informed decision about which of either the T-SES-TALA or the NC-SES-TALA performs better according to the parameters produced.

Table 5.8: Screen shot of items for the NC-SES-TALA part 1 & 2 in numbers

Memo	4	5	1	3	2	4	3	3	4	4	3	4	3	3		3	3	4
Item number	1	2	3	4	5	6	7	8	9	10	11	12	13	14		15	16	17
Student No.																		
1	4	1	3	2	2	3	2	1	3	4	1	3	1	1		2	3	1
2	2	1	2	1	2	2	3	1	3	4	2	2	1	3		3	2	1
3	4	1	2	3	5	4	3	3	4	4	3	2	3	2		1	3	1
4	5	1	2	3	4	3	3	3	4	4	3	2	3	3		3	3	4
5	2	5	4	1	3	3	3	3	4	4	3	2	2	3		3	3	4

5.4 Administering the test

This section deals with how both the T-SES-TALA parts 1&2 and NC-SES-TALA parts 1&2 were administered at the four different schools where the experiment was performed. In order to ensure that conditions used for this test were conducive for the administering of the test, the test was administered for 55 minutes with hard copies that were the same in all respects. It was also administered under more or less the same weather conditions that were not harsh for learners. It was also ensured that ventilation at venues was conducive for air flow. Venues used for the Grade 12 NSC were the same venues that were used for these tests. Some schools made use of several venues while others had one big venue for all the participants. Schools where tests were administered had promised to supply teachers who would help with invigilation during testing. However, as a measure of precaution, the researcher took along an additional two assistants to school centres to mediate problems which might ensue every time testing was to take place. For this reason, enough personnel were at hand for the test sessions for this study. As a first step, each individual learner was

issued with a pencil, eraser, and an answer sheet. A second step was to explain details pertaining to the test and the manner in which the answer sheet was to be filled out. As a third step, the question paper was issued after which the instructions written on the cover of the question paper were discussed with learners. Learners were informed that the question paper was scheduled for 55 minutes after which no further writing was expected. The same routine was followed for the administering of each question paper for each session. Each session involved the writing of both part one and part two which brought the writing of both parts to approximately two hours. There was a break of 15 minutes between each session. The interval between the Translated TALA and the Newly Created TALA varied between a minimum of five days and a maximum of 12 days as can be seen above in the section on “selecting participants”. Although the method used for this interval was not meant to be that of a test-retest procedure, the cue was taken from this that the interval should not be too long (more than six months) and should always be reported (Weir, 2005a:25).

5.5 Collection and safe keeping of stimulus material

Every question paper and answer sheet used during the administration of the tests was collected immediately after the test. The question papers that were used were stored in offices at the North-West University (NWU), and will be destroyed after a period of five years of safe keeping according to NWU record-keeping guidelines. No learners were allowed to leave the venue until all question papers and answer sheets as well as stationery were collected. This was to ensure that no papers were left the venue to prevent learners from using such papers for any other purpose. Even scrap answer sheets were not left behind but disposed of at the NWU.

5.6 Methods and measures for TALA validation

It was stated in the introduction to this chapter that this study follows a quantitative research design. It was also stated that decisions that would be taken at the conclusion of this study would be based on statistical analysis in order to investigate the properties of the data that would be generated.

Measures that were produced by both *TiaPlus* and *Winsteps* were defined and explained using tables and graphs. These tables and graphs were used to clarify tendencies and trends that are important for conclusions and forecasts that can be drawn from this experiment. These conclusions and forecasts are beneficial for the framing of the validity argument for this study. This validity argument will be able to inform the researcher which of the two tests, i.e. the T-SES-TALA and the NC-SES-TALA, is best suited to serve as a replacement section of 20% in the Grade 12 Sesotho HL examination paper.

5.7 Conclusion

This chapter outlined issues that were important concerning the method that was followed when the empirical investigation was done for this study. It demonstrated steps that were followed when permission was sought from authoritative bodies and institutions in line with requirements for ethical standards. It also outlined how the stimulus material that was used in this study was developed. This included the translation of the original English TALA into the Translated Sesotho TALA according to language testing requirements. In addition to this, it discussed the process used in the development of the Newly Created Sesotho TALA with regard to meeting theoretical standards for language test development. It was also able to outline how the participants who were used in this experiment were selected in accordance with ethical standards for selecting participants for an experimental study of this nature. Lastly, it described how the test was administered and how data that was collected for this study would be analysed.

CHAPTER 6: EMPIRICAL RESEARCH RESULTS

6.1 Introduction

It was explained in Chapter 1 that a specific procedure would be followed in obtaining quantitative results from the data generated by this study and that these results will be used for a validity argument. The results are obtained from the analyses of a number of data sets that were discussed in Chapter 5.

The research question for this study as stated in Chapter 1 is which test between the Translated Sesotho TALA and the Newly Created Sesotho TALA shows a higher level of equivalence with the English TALA when investigated empirically? Consequently, the best performing TALA shall be recommended to form 20% of the 300 marks for the Grade 12 home language examination paper because of its higher validity. Making use of the *TiaPlus* and *Winsteps* statistical programs for data analysis should enable one to determine which of the two tests (i.e., the T-SES-TALA and the NC-SES-TALA) is more appropriate and adequate for the purposes of this study. In other words, a comparison of the parameters obtained for the T-SES-TALA and the NC-SES-TALA should provide the researcher with the statistical evidence necessary to determine which of the two tests has the best internal and external validity, and as a consequence, which of the two tests are most adequate and appropriate for consideration to form part of the Grade 12 HL examination paper.

The hypothesis for this study created the need to answer specific research questions which are mentioned again below and can only be affirmed by the analysis of the data sets that were discussed in Chapter 5:

Which test between the Translated Sesotho TALA and the Newly Created Sesotho TALA shows a higher level of equivalence with the English TALA when investigated empirically?

In order to address this question, the following sub-questions need to be answered:

1.1 When considering both the Translated Sesotho TALA and the Newly Created Sesotho TALA, which of these two TALAs shows better evidence of internal validity with the English TALA, with regard to:

- construct validity;
- face validity; and
- content validity?

1.2 When considering both the Translated Sesotho TALA and the Newly Created Sesotho TALA, which of these two TALAs shows better evidence of external validity with the English TALA with regard to:

- predictive validity;
- concurrent validity; and
- consequential validity.

6.2 Data analysis

The language test analysis software programs that were used for data analysis in this study were used to reveal the parameters expected of language tests such as the T-SES-TALA and the NC-SES-TALA. The parameters referred to here include item-biserial correlation and facility indices.

6.2.1 Raw scores

The marking of the score sheets was done electronically by a marking scanner (for both the 83 and the 104 items of the two tests, i.e., the T-SES-TALA part 1 & 2 and NC-SES-TALA part 1 & 2). The relevant information from these score sheets that was captured in a spreadsheet format included details, such as student number, marks obtained by each candidate, memorandum for each item, weight for each question, number of students, average class mark, and average class percentage.

In order to create a data set from these score sheets for the T-SES-TALA, the 83 items for T-SES-TALA part 1 and the 104 items for T-SES-TALA part 2 for all the schools were combined into one data set. A similar approach was followed in order to create a second data set for NC-SES-TALA, i.e. the 83 items for NC-SES-TALA part 1 and

the 104 items for NC-SES-TALA part 2 for all the schools were combined into one data set. Each data set was made up of 187 items overall as can be seen from Tables 6.1 and 6.2.

Table 6.1: Total number of items for T-SES-TALA part 1 & 2 (1st data set)

SCHOOL	TALA TYPE	NUMBER OF ITEMS
1-4	T-SES-TALA part 1	83
1-4	T-SES-TALA part 2	104
	TOTAL	187 per test

Table 6.2: Total number of items for NC-SES-TALA part 1 & 2 (2nd data set)

SCHOOL	TALA TYPE	NUMBER OF ITEMS
1-4	NC-SES-TALA part 1	83
1-4	NC-SES-TALA part 2	104
	TOTAL	187 per test

The two data sets were entered into the *TiaPlus* and *Winsteps* language testing software programs for computation to determine the parameters of each test. *TiaPlus* was selected because it is ideal for computing item point-biserial correlation and facility indices which are used to determine the performance of individual test items, the entirety of the test, and the relation between different subtests (Steyn, 2018:67). *TiaPlus* uses both CTT analysis and Rasch analysis to produce parameters for both internal and external validities. The benefit for making use of the *TiaPlus* software is that it provides both descriptive and inferential statistics about tests. *Winsteps* was used to augment the *TiaPlus* analysis so as to verify the parameters that were produced by *TiaPlus*.

6.2.2 First view of the analysed scores

The overview of results that were generated first by *TiaPlus* for the T-SES-TALA part 1&2 was disappointing and needed closer examination (see Tables 6.3–6.5 taken from the *TiaPlus* analysis of results). Notwithstanding many possible variables influencing reliability, these results indicated that, in terms of the reliability (alpha) value, the T-SES-TALA part 1&2 had a lower reliability of 0,41 compared to that of the ENG-TALA and the NC-SES-TALA part 1&2 which stood at 0,96 and 0,82 respectively. The intercorrelation between subtests of the T-SES-TALA were found to be negative and weak, which meant that the reliability between these subtests was low. This lack of strength regarding the correlation between subtests also indicated that it did not render expected results regarding differentiation, i.e. items did not discriminate well between strong and weak learners when one compares the performance of learners according to subtests. It was further found that for the translated test, the average test score of 50,53 was lower when compared to the ENG-TALA and the NC-SES-TALA part 1&2 that produced average test scores of 116,79 and 52,20 respectively (this is slightly lower when compared to that of the T-SES-TALA) out of 187 respectively. Possible reasons for this poor performance by the T-SES-TALA are discussed in Section 6.2.3 under “detailed analysis of the data” below. The frequency histogram (cf. Figure 6.1) was also skewed to the left, indicating that fewer learners got answers correct. The factor analysis (Figure 6.2) further showed many outliers, indicating that many test items measured a second factor (i.e., another domain than the one tested). The factor analysis (generated by *TiaPlus* based on the item intercorrelation) in Figure 6.2 includes the six different subtests that form part of the 187 items listed in the scatterplot – all these form part of the single domain being tested. The six subtests illustrated in this scatterplot are the two scrambled texts, text comprehension, interpreting graphs and visual information, vocabulary knowledge, and grammar and text relations. As their names indicate, the subtests measured different factors. The purpose and factors that each each of these subtests measured were explained in Section 5.3 under “preparation of the stimulus material”. The scatterplot shows 78 different outliers indicating that it measures other factors than those intended for the six subtests

mentioned above. This overview was necessary in order to guide actions that are appropriate for further exploration of the data without which a final conclusion about the performance of the tests would have been premature.

Table 6.3: Key parameters of T-SES-TALA part 1&2

SubGroup number	: 0	SubTest number	: 0
Number of persons in test	: 514	Number of selected items	: 187
Minimum test score	: 0	Maximum test score	: 187
Average test score	: 50,53	Standard deviation	: 7,72
Average P-value	: 27,02	Std. Error of Measurement	: 5,94
Average Rit	: 0,09		
Coefficient Alpha	: 0,41	SE Coeff. Alpha	: 0,04
GLB	: 0,00	Asymptotic GLB coef	: 0,00
Items used in GLB proc	: 187		
Cut-off score	: 60,5	Percentage failing	: 91,05

Table 6.4: Key parameters of NC-SES-TALA part 1&2

SubGroup number	: 0	SubTest number	: 0
Number of persons in test	: 519	Number of selected items	: 187
Minimum test score	: 0	Maximum test score	: 187
Average test score	: 52.20	Standard deviation	: 13.97
Average P-value	: 27.92	Std. Error of Measurement	: 5.95
Coefficient Alpha	: 0.82	SE Coeff. Alpha	: 0.00
Average Rit	: 0.17		
GLB	: 0.00	Asymptotic GLB coef	: 0.00
Items used in GLB proc	: 187		
Cut-off score	: 60.5	Percentage failing	: 79.58
Misclassifications:			
Alpha based		GLB based	

Table 6.5: Key parameters of ENG-TALA

Subgroup number	:0	Subtest number	:0
Number of persons in test	: 690	Number of selected items	: 187
Minimum test score	:0	Maximum test score	:187
Average test score	: 116.79	Standard deviation	:25.19
Average P-value	: 62.45	Std. Error of Measurement	: 5.64
Coefficient Alpha	: 0.95	SE Coeff. Alpha	:0.00
GLB	: 0.00	Asymptotic GLB coef	:0.00
Items used in GLB proc	:187		

Cut-off score : 61.5
:1.45

Percentage failing

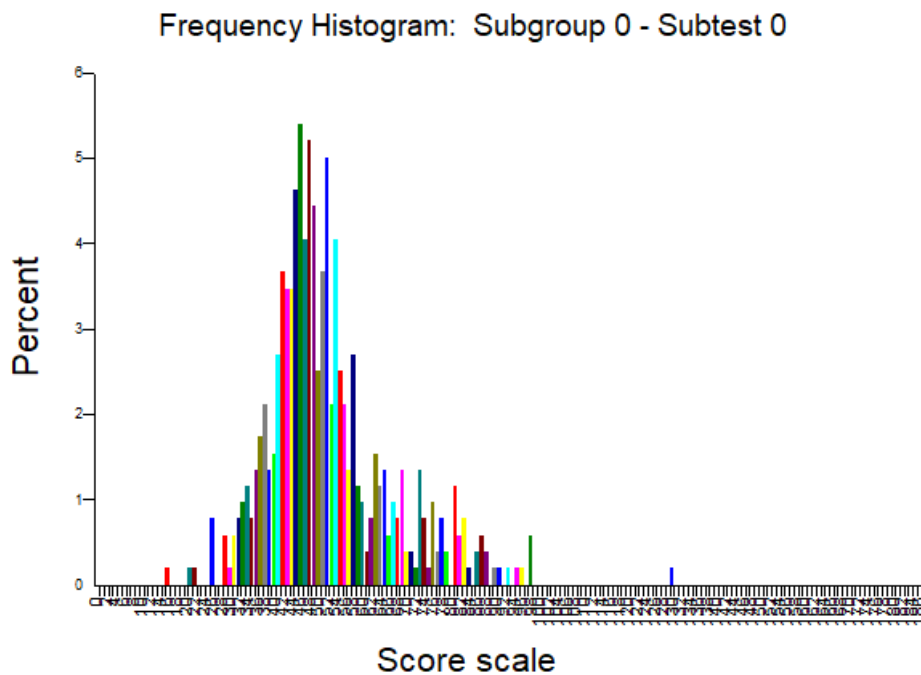


Figure 6.1: Frequency histogram: T-SES-TALA subgroup 0-Subtest 0

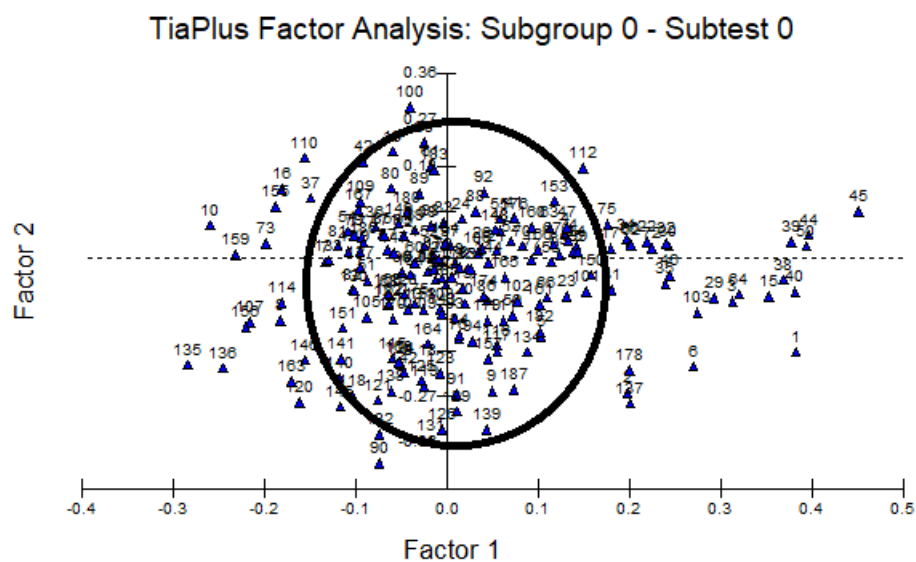


Figure 6.2: *TiaPlus* factor analysis for T-SES-TALA part 1&2

As indicated in the overview above, the preliminary results for T-SES-TALA part 1&2 specifically necessitated further analysis of the data in order to explore issues that could shed light on the low reliability measures as well as problems indicated in the factor analysis. At this point, it was speculated that the reason for the low performance of T-SES-TALA part 1&2 may be found in a possibly faulty memorandum, patterned answers provided by learners (answers that formed specific patterns, for instance A, B, C, D and then again A, B, C, D which often include guessed answers), and/or incomplete answer sheets. Therefore, the memorandum was checked for any faulty answers, all patterned answers were scrutinised, and incomplete answer sheets were also inspected. The results showed that the memorandum was in fact correct; all answer sheets that showed clear patterned answers were removed from the data set, and all incomplete answer sheets were eliminated from the analysis. This process of removing spoiled answer sheets from the tally, was followed for both Sesotho tests (i.e., for both Sesotho TALAs - part 1 or 2) and for all schools (i.e., schools 1, 2, 3, and 4). An average of 21,5 answer sheets were removed from each school. However, even after this cleaning up of the data, this test still failed to perform adequately: its overall reliability went down from 0,41 to 0,33. The overall low performance of the T-SES-TALA led to a decision that – in the interest of responsible design – no further analysis was necessary since this test clearly failed to meet the basic requirements of overall reliability, especially when it is compared with the NC-SES-TALA. However, some of the statistical information related to the T-SES-TALA would be used for further explanation in the remainder of this chapter. A conclusive decision about the NC-SES-TALA could, therefore, not be taken until further investigations and comparisons were completed.

The focus of analysis accordingly shifted to the Newly Created Sesotho TALA (NC-SES-TALA). The preliminary results for this test had shown promising parameters with regard to discrimination index, facility value, and overall reliability level. Its overall reliability coefficient (an alpha value of 0,82) was considerably higher than that of the T-SES-TALA. It further had fewer outliers when compared to those of T-SES-TALA identified from its scatterplot (78 in number). However, in order to make any conclusive

statement about this test, additional parameters such as the intercorrelation between subtests, and fairness had to be explored. It must be emphasised here that although the outliers were fewer, factor analysis was still problematic as these outliers were again measuring a second factor/domain, which was not ideal. It should also be indicated that from the preliminary results it was evident that the test had a high difficulty level, indicated by the low average score of 27,92% achieved by the learners (note that *TiaPlus* calculates this in a percentage score and not as a raw score). Therefore, a closer analysis of the results was necessary to ensure that the test met several critical requirements for an adequate and appropriate test. Certain additional critical parameters for the NC-SES-TALA needed to be systematically examined as well, if an informed decision was to be reached about this test. These included a focus on both the item facility values (difficulty level) of both the overall test and subtests as well as factor analyses of the overall test and subtests. This is important because both item facility and discrimination values indicate the adequacy of a test (Van der Walt & Steyn, 2008:195). These were done using both Classical Test Theory (CTT) and IRT. As indicated, *TiaPlus* has the capacity to generate important parameters pertaining to both CTT and Rasch analysis. Additionally, a statistical report by *Winsteps* will be presented in order to corroborate the results from the *TiaPlus* analysis.

6.2.3 Detailed analysis of the data

For purposes of performing the detailed analyses referred to above, CTT and IRT analyses using the *TiaPlus* and *Winsteps* statistical programs were used. The *Winsteps* statistical program was employed to augment the *TiaPlus* analysis in order to inform judgement about the productivity of the items belonging to both the T-SES-TALA and the NC-SES-TALA. The *TiaPlus* program uses Rasch principles to provide values for the performance of test items. *Winsteps* produces what is referred to as convergence tables to display the performance of items and persons, as well as comparisons between subtests (cf. Tables 6.6, 6.12, 6.13, 6.16, 6.19, 6.23 and 6.25). It describes the performance of items in terms of item reliability and person reliability. These tables are necessary because they provide a visual table to discern how items match up against test takers. This statistical program has the advantage of helping

test developers determine important parameters such as the difficulty level of items and person reliability measured very quickly (Green, 2013:167). Since person reliability is a different concept from the usual reliability in which the performance of items is described, person reliability needs to be clarified. Person reliability is defined as a measure that helps test designers better understand the degree to which persons (test takers) fit the model used by *Winsteps* to analyse tests by means of a statistic called the Infit Mean Square and Outfit Mean Square. The concept of 'model' as defined by *Winsteps* refers to a built-in item and person response model that is used by this program to produce convergence tables which provide information about how well the data fit this model and the degree of confidence one can have in the measurements produced. Two ways in which this is achieved is by providing two measures of reliability, namely, for specific items and persons. This way people can discern at a glance whether the items have given an accurate picture of test takers' abilities and consequently, that the test takers have provided a true picture of the level of difficulty of items (Green, 2013:167).

The parameters expected for the NC-SES-TALA as the chosen test for further refinement are based on other tests of academic literacy belonging to ICELDA as stated in paragraph 6 below in this section. These parameters were also endorsed by the project team for the development of the TALA from the onset of the project (Steyn, 2018:57). These parameters cannot deviate from the ones that were set by the project team as they serve as a criterion for all Grade 12 HL TALA tests that will be developed for other South African official languages in future, if endorsed by the DBE. Making use of these parameters, the following questions had to be answered based on the statistics produced by *TiaPlus* and *Winsteps* using either CTT or Rasch principles:

1. Do the items have an item point biserial of above 0,2 or discrimination index of greater than 0,25?
2. Is the facility value of items above 0,15 and below 0,84?
3. Is the overall reliability level at 0,9 or above?
4. Are the subtest intercorrelations between 0,2 and 0,5?

Using *TiaPlus*, the first three questions can be answered based on CTT principles. The answer to the fourth question will be determined by the results of the Rasch analysis, using *Winsteps*. While only pertinent statistics will be presented and discussed in the text below, a comprehensive statistics report can be found in appendices A1 to A3 of this thesis.

For question 1 above, the expectation is that, for items to be able to discriminate well between higher and lower ability learners, the Pearson point biserial (rpbis) should be above 0,2 (Steyn, 2018: 60). It should be noted that because *TiaPlus* uses the item-total correlation (Rit) (which is also known as the discrimination index) to describe the rpbis, it can range between -1,0 and 1,0 (Australian Medical Assessment Collaboration, not dated: 24). The Rit for the NC-SES-TALA in this study displays an average of 0,17, which is not an acceptable rpbis for a test of this nature because it is below 0,2 (see Table 6.4). This is not in agreement with the rpbis of the ENG-TALA which scored 0,27 (Steyn, 2018:75) (see Table 6.5). This score, although disheartening, is better than the T-SES-TALA which shows an average Rit of 0,09 which, when rounded up, only reaches 0,1 (see Table 6.3). It can therefore be concluded that NC-SES-TALA discriminates slightly better between higher and lower ability learners when compared with the T-SES-TALA (while still not at an acceptable level).

For the second question, values between 0,15 and 0,84 are the acceptable average P-values for the parameters of the whole test. These P-values are in line with facility values which the Test of Academic Literacy Levels (TALL) and its Afrikaans counterpart, Toets van Akademiese Geletterdheidsvlakke (TAG) are prepared to accept (Weideman, 2011:105). These parameters are thus acceptable for the TALA because of the test being modelled on these two tests. The parameters are also those accepted by the project team to which this study belongs. The P-value stands for 'percentage correct' and it represents the proportion of candidates who answered the specific item correctly. As stated in the question, for the purpose of this study, the P-value should be above 0,15 and below 0,84. This implies that items that are too easy have a P-value of 0,84 and above (i.e., where 84% and more of the candidates

answered the item correctly), and those that are too difficult with a P-value of 0,15 and below (where less than 15% of candidates got them correct), should be avoided. This parameter is crucial in cases where a test has to be refined (which will also be done later on in the chapter for this specific test). However, based on the available scores (cf. Table 6.3), it is obvious that the T-SES-TALA was more difficult than the NC-SES-TALA, albeit by a small margin. The average P-value for the T-SES-TALA is 0,27 while for the NC-SES-TALA it is 0,28. These values are much lower than that of the ENG-TALA which is 0,62. In the final analysis of the P value, it can therefore be concluded that more learners from the NC-SES-TALA got items correct as compared to the number of learners from the T-SES-TALA who got items correct. This is also highlighted by the fact that the percentage of test takers who failed the T-SES-TALA is 91,05% and 79,58% for the NC-SES-TALA (cf. Table 6.4 and 6.5).

For the third question, *TiaPlus* calculated the overall Cronbach alpha for the NC-SES-TALA as 0.82. This indicates that it is substantially more reliable than the T-SES-TALA which scored 0.41 (cf. Tables 6.3 – 6.5). Based on the fact that the groups were homogeneous, although not exactly the same size because of the participants that were flagged and removed as well as the number of questions that were exactly the same in the two tests but did not discriminate equally, it can be asserted at this point that the NC-SES-TALA is more reliable than the T-SES-TALA depending on this parameter. This is one of the reasons why the NC-SES-TALA (and not the T-SES-TALA) was considered for further refinement and development.

Moreover, when one considers the analysis from *Winsteps* (see convergence Tables 6.6 and 6.7), the overall person reliability and item reliability for the T-SES-TALA are ,44 and ,94 respectively, while the overall person reliability and item reliability for the NC-SES-TALA are ,80 and ,95 respectively. These figures imply that the T-SES-TALA did not perform as well as the NC-SES-TALA in terms of how test items and persons had performed in these two tests. The ,44 person reliability figure in the T-SES-TALA shows the extent to which the number of persons would be replicated if this sample of persons were to be given another set of items measuring the same construct (Green, 2013:159). On the other hand, the ,80 reliability figure in the NC-SES-TALA shows the

extent to which this number of persons would be replicated if this sample of persons were given a set of items measuring the same construct. Just like in other types of reliability, the higher and more positive the figure, the better. In this instance, therefore, it is clear that the NC-SES-TALA has a higher person reliability than the T-SES-TALA. In the same way as for CTT, this figure is affected by factors including, but not limited to, test takers and items involved, item distributions and their appropriateness to each other, motivation, and emotional state of test takers. Also, when one compares the item reliability of ,94 for the T-SES-TALA to that of the NC-SES-TALA of ,95 in the *Winsteps* convergence table (on a scale of -1 to +1), one sees that the reliability of the NC-SES-TALA is slightly higher than that of the T-SES-TALA. The person reliability figures provide an estimate of the confidence one can have in the test takers' results, i.e., the likelihood of how these persons may perform on another set of items of a similar nature given under similar conditions. Moreover, the item reliability figures provide an estimate of the confidence one can have in the performance of the items with another test population given under similar conditions. Just like in CTT, the higher the reliability level, the better confidence we are likely to have in the accuracy of the measures given. This is in agreement with statistics shown by *TiaPlus*. However, because of the overall person reliability, which is lower than the required measure (the required measure should be at least ,85 and not ,80), the NC-SES-TALA test needed further refinement.

Table 6.6: T-SES-TALA 187 Item convergence table

```
513 PERSON Records Input
Building Index ...
      CONVERGENCE TABLE
-Control: \ice02666_Sesotho1_187item.txt Output: \ICELDA\ZOU334WS.TXT
|  PROX      ACTIVE COUNT      EXTREME 5 RANGE      MAX LOGIT CHANGE |
|  ITERATION PERSON  ITEM CATS      PERSON  ITEM      MEASURES STRUCTURE|
>=====<
|    1    513    187    2    1.41  2.11      -2.0149      |
Index Size: 45 with 1 Patterns
>=====<
|    2    513    187    2    1.46  2.13      -.0718      |
Probing data connection: to skip out: Ctrl+F - to bypass: subset=no
Processing unanchored persons ...
>=====<
Data fully connected. No subsets found
|Control: \ice02666_Sesotho1_187item.txt Output: \ICELDA\ZOU334WS.TXT
|  JMLE      MAX SCORE  MAX LOGIT      LEAST CONVERGED  CATEGORY STRUCTURE|
```

```

| ITERATION  RESIDUAL*  CHANGE PERSON  ITEM  CAT  RESIDUAL  CHANGE|
>=====<
| 1      -.70   -.0049   15   13*      |
-----
Calculating JMLE Fit Statistics
>=====<
Time for estimation: 0:0:1.118
Output to C:\Albert Documents\ICELDA\ZOU334WS.TXT
ice02665_TALA_Sesotho1_187_item.xlsx
-----
| PERSON  513 INPUT  513 MEASURED      INFIT      OUTFIT  |
|  TOTAL    COUNT  MEASURE REALSE    IMNSQ  ZSTD OMNSQ  ZSTD|
| MEAN   50.5   187.0   -1.04   .17   1.00   .0  1.00   .0|
| P.SD    7.7    .0    .23   .01   .04   .5  .06   .6|
| REAL RMSE .17 TRUE SD .15 SEPARATION .88 PERSON RELIABILITY .44|
|-----|
| ITEM   187 INPUT  187 MEASURED      INFIT      OUTFIT  |
|  TOTAL    COUNT  MEASURE REALSE    IMNSQ  ZSTD OMNSQ  ZSTD|
| MEAN  138.6   513.0    .00   .10   1.00   .0  1.00   .0|
| P.SD   44.5    .0    .44   .01   .01   .4  .02   .5|
| REAL RMSE .10 TRUE SD .43 SEPARATION 4.14 ITEM RELIABILITY .94|
|-----|
Output written to C:\Albert Documents\ICELDA\ZOU334WS.TXT
CODES= 01
Measures constructed: use "Diagnosis" and "Output Tables" menus

```

Table 6.7: NC-SES-TALA 187 Item convergence table

```

518 PERSON Records Input
Building Index ...
      CONVERGENCE TABLE
-Control: ice02654_Sesotho2_Winsteps.txt Output: \ICELDA\ZOU425WS.TXT
| PROX      ACTIVE COUNT      EXTREME 5 RANGE      MAX LOGIT CHANGE |
| ITERATION PERSON  ITEM CATS   PERSON  ITEM   MEASURES STRUCTURE|
>=====<
| 1   518   187   2    2.25  2.07   -2.3691      |
Index Size: 69 with 1 Patterns
>=====<
| 2   518   187   2    2.35  2.12   -.0976      |
Probing data connection: to skip out: Ctrl+F - to bypass: subset=no
Processing unanchored persons ...
>=====<
Data fully connected. No subsets found
|Control: ice02654_Sesotho2_Winsteps.txt Output: \ICELDA\ZOU425WS.TXT
| JMLE  MAX SCORE  MAX LOGIT  LEAST CONVERGED  CATEGORY STRUCTURE|
| ITERATION  RESIDUAL*  CHANGE PERSON  ITEM  CAT  RESIDUAL  CHANGE|
>=====<
| 1      -1.30   -.0090   68   80*      |
>=====<
| 2      -.42   -.0030   68   80*      |
-----
Calculating JMLE Fit Statistics
>=====<
Time for estimation: 0:0:1.975
Output to C:\Albert Documents\ICELDA\ZOU425WS.TXT
ice02653_Sesotho2_Winsteps.xlsx
-----
| PERSON  518 INPUT  518 MEASURED      INFIT      OUTFIT  |
|  TOTAL    COUNT  MEASURE REALSE    IMNSQ  ZSTD OMNSQ  ZSTD|
| MEAN   52.2   187.0   -1.03   .17   1.00  -.1  1.00   .0|

```

```

| P.SD   14.0    .0    .38    .01    .04    .7    .08    .8|
| REAL RMSE .17 TRUE SD .34 SEPARATION 1.98 PERSON RELIABILITY .80|
|-----|
| ITEM   187 INPUT   187 MEASURED          INFIT      OUTFIT |
|      TOTAL    COUNT  MEASURE REALSE    IMNSQ  ZSTD OMNSQ  ZSTD|
| MEAN   144.6   518.0    .00   .10    1.00  -.1  1.00  -.1|
| P.SD   44.3    .0    .49    .01    .04  1.3  .05  1.3|
| REAL RMSE .11 TRUE SD .47 SEPARATION 4.51 ITEM RELIABILITY .95|
|-----|

Output written to C:\Albert Documents\ICELDA\ZOU425WS.TXT
CODES= 01
Measures constructed: use "Diagnosis" and "Output Tables" menus
Processing Table 11
Sorting 187      by Fit for Table 11.1
>=====<
Processing Misfitting ITEM for Table 11
>=====<

```

For the fourth question mentioned in Section 6.2.3, the *TiaPlus* report indicates the correlation between subtests. The five subtests discussed here are: scrambled text, text comprehension, graphs and visual information, vocabulary knowledge, and grammar and text relations. As stated in the question, the intercorrelation between subtests should, according to the parameters that they measure, be between 0,2 and 0,5. It is also noteworthy to observe that these parameters conform to those accepted by other expert organisations and individual scholars. For instance, Alderson *et al.* (2005:184), posit that the correlation between subsections should be fairly low – they should range between 0,3 to 0,5. According to them, if two sections correlate too highly, it might be possible that they are essentially testing the same attribute on the one hand (see also Van der Walt & Steyn, 2007). On the other hand, if they test too low, it might be the case that it measures a factor that lies outside of the domain being tested. The following Tables (6.8 and 6.9) taken from the *TiaPlus* report indicate the difference in measurements between the subtests of the T-SES-TALA and the NC-SES-TALA.

Table 6.8: T-SES-TALA subtest intercorrelations

Table of Subtest Intercorrelations

	Total	Subtest(s)						
Subtest	test	1	2	3	4	5	6	
Scrambled text	1	0,20						
Scrambled text	2	0,20	0,06					
Text comprehens	3	0,57	0,02	0,00				
Interpreting gr	4	0,44	0,05	0,05	0,01			
Vocabulary know	5	0,54	0,03	0,10	-0,05	0,11		
Grammar & text	6	0,50	-0,03	0,00	-0,08	0,12	0,18	

Table 6.9: NC-SES-TALA subtest intercorrelations

Table of Subtest Intercorrelations

	Total	Subtest(s)						
Subtest	test	1	2	3	4	5	6	
Scrambled text	1	0,21						
Scrambled text	2	0,16	0,07					
Text comprehens	3	0,82	0,18	0,15				
Interpreting gr	4	0,42	0,05	-0,03	0,08			
Vocabulary know	5	0,58	-0,02	-0,01	0,16	0,37		
Grammar & text	6	0,40	-0,02	-0,06	-0,01	0,19	0,30	

Based on Table 6.8, the majority of subtests with regard to intercorrelations between different subtests did not perform well and only one subtest intercorrelation (marked in green in the table) nearly meet the minimum subtest intercorrelation requirement of 0,2 for the T-SES-TALA. However, it is clear from Table 6.9 depicting the results for the NC-SES-TALA that six subtest intercorrelations (also marked in green in the table) meet the subtest intercorrelation of between 0,2 and 0,5. This is an indication that different subtests measured different attributes for NC-SES-TALA.

The above findings with regard to the comparison between the two Sesotho TALA tests (with specific reference to the subtest intercorrelations) together with the correlation between subtests and the whole tests will be used in the formulation of a validity argument (cf. Van der Walt & Steyn, 2007) later on in the chapter. However,

before embarking upon outlining the validity argument for this study, it is important to explain the route to be taken here after.

A number of salient properties of the T-SES-TALA and the NC-SES-TALA have been highlighted in the analysis above. At this stage of the analysis, these properties include: the substantial difference between the overall reliability coefficient of the T-SES-TALA and the NC-SES-TALA (in favour of the NC-SES-TALA); the difference between measures of ability where the NC-SES-TALA discriminates substantially better between items than the T-SES-TALA; the noteworthy difference in subtest intercorrelation between the T-SES-TALA and the NC-SES-TALA (that favours the NC-SES-TALA); and lastly, the better facility value of the NC-SES-TALA when compared to that of the T-SES-TALA (indicating that the T-SES-TALA was more difficult than the NC-SES-TALA). Based on these properties, a decision was taken to subject the NC-SES-TALA to further statistical analysis (using acceptable language testing measures) in an attempt to improve the performance of the NC-SES-TALA. It was therefore decided that the performance of the T-SES-TALA did not warrant further development. The decision taken was, therefore, to refine the NC-SES-TALA, based on further statistical analysis, to determine if it could be further improved.

6.2.4 NC-SES-TALA refinement

The refinement that was done for the NC-SES-TALA was directed towards selecting those test items with better characteristics from the 187 items of the test. In other words, items with stronger attributes, and therefore, better productivity (items that discriminated better and had better facility values) could be included towards a potentially better overall performance of the test. This would help the researcher to formulate a stronger validity argument and also explain why this test may be preferred in terms of performance rather than the T-SES-TALA.

For the process of refining the NC-SES-TALA, a team of three experts was convened. These experts have extensive knowledge about the development of tests of academic literacy in one form or the other. For instance, Albert Weideman – a professor of

Applied Language Studies and Research Fellow at the University of the Free State – is a well-known expert in test design. The other expert, Gustav Butler – an associate professor and Director of Research at a North-West University research entity called Understanding and Processing Language in Complex Settings (UPSET) - has experience in the translation of the Test of Academic Literacy Levels (TALL) into Sesotho. Lastly, Tobie Van Dyk is a professor of Applied Linguistics and Academic Literacy at the North-West University and has extensive knowledge about language testing.

For the refinement of the test, 60 items from a total of 187 items were to be selected with the aim of producing a reduced, more reliable and valid test. While refinement of tests is a normal activity in various disciplines, this line of test refinement has often been followed for similar tests in ICELDA-led test development projects. For instance, a refinement of the Test of Academic Literacy for Postgraduate Students was completed by Rambiritch (2013) in her study “Validating the Test of Academic Literacy for Postgraduate Students”. This specific test was refined three times before a final version was accepted by the researcher (Rambiritch, 2013). Another test worth mentioning for which the same procedure of refinement was used is the ENG-TALA that was developed by Steyn (2018) as part of her master’s dissertation titled, “A theoretical justification for the design and refinement of a Test of Advanced Language Ability (TALA)”. For this test, 187 items were reduced to 60 items. As discussed before, this specific test was used as a model for the two Sesotho tests that are the focus of this study. Although only two examples are provided here, a number of other ICELDA-led tests have undergone the same procedure of refinement. Refining newly designed tests is beneficial in that it reveals aspects of the test that the designer should consider in a decision about whether such a test should be further developed. Piloting and re-piloting of a test is, therefore, an important step in test design within this context. For this refinement, the team of experts referred to above, focused on the selection of items based on various attributes of items that could be used to improve the parameters of the NC-SES-TALA so that only 60 items from 187 items with better

attributes could remain. In turn, these 60 items would be subjected to a last and final statistical analysis.

From the 187 item test, only 5 items from 10 items in the scrambled text section were supposed to be retained, 20 items from 73 text comprehension items, 8 items from 30 interpreting graphs and visual information items, 15 items from 32 vocabulary items and 12 from 42 grammar and text relations items. All retained items would leave 60 items in total when added together (i.e., $5+20+8+15+12 = 60$), which will resemble the test structure of the refined English TALA. It was important to divide the different TALA sections according to these numbers in order to correlate with the specifications of the original English TALA. These specifications refer to the number of items which are supposed to be contained in each section of the test and the details on different skills which each section sought to assess, as well as the number of items which are supposed to be allocated to each skill. This would enable concurrent comparison between the different TALAs. It should also be remembered that refining the 187 TALA into 60 items is in line with the theoretical assertion that was advanced in this study that TALA should replace 20% of the 300 marks for Grade 12 HL examination papers.

The basic point of scrutiny for the various subtests and test items by the panel was the item productivity based on the principles of CTT and Rasch analysis – i.e., the P-values and the Rit (discrimination index) values for each item as produced by the *TiaPlus* statistical program. This was augmented by additional information derived from the statistical analysis using the *Winsteps* statistical program. Additional explanations from statistics from *Winsteps* were added to explanations on statistics from *TiaPlus* at places in the course of this discourse if necessary. In order to identify items that displayed better P and Rit values, subtests and items from the *TiaPlus* statistical analysis were systematically evaluated. Excerpts of both the T-SES-TALA and the NC-SES-TALA analyses, displayed in Tables 6.10 and 6.11 and taken from the *TiaPlus* report, indicate performance differences (in terms of P and Rit values) between individual items belonging to the two Sesotho TALA tests, first for the scrambled texts and later for the other subtests.

Table 6.10: T-SES-TALA *TiaPlus* report for the items making up subtest 1 and 2

Item	Item	----- P- and A- values -----										Mis-	----- Weighted -----									
Label	nr.	Weight	Key	A	B	C	D	E	F	O/D	sing	Max	Mean	P	Sd	RSK	[Lo	Rit	Hi]	Rir	AR	
1	1	C	10	7	59*	8	13			3	14		1	0,59	59	0,49	0,49	6	15	23	8	40
2	1	A	28*	17	13	12	25			4	19		1	0,28	28	0,45	0,45	-1	9	18	3	41
3	1	E	18	34	12	18	14*			4	20		1	0,14	14	0,35	0,35	5	14	22	9	40
4	1	B	17	28*	6	25	20			4	23		1	0,28	28	0,45	0,45	-2	8	14	0	41
5	1	D	26	12	8	34*	15			5	27		1	0,34	34	0,47	0,47	5	13	22	7	40
6	1	D	5	25	8	53*	6			3	16		1	0,53	53	0,50	0,50	10	19	27	12	40
7	1	E	26	18	27	13	13*			3	16		1	0,13	13	0,34	0,34	-1	7	15	3	41
8	1	C	25	20	20*	11	20			4	22		1	0,20	20	0,40	0,40	-2	6	15	1	41
9	1	B	26	14*	21	11	24			4	20		1	0,14	14	0,35	0,35	6	14	23	10	40
10	1	A	17*	23	24	8	24			4	20		1	0,17	17	0,37	0,37	-11	-1	8	-6	41

Table 6.11: NC-SES-TALA *TiaPlus* report for the items making up subtest 1 and 2

Test and Item Analysis

Item	Item	----- P- and A- values -----										Mis-	----- Weighted -----									
Label	nr.	Weight	Key	A	B	C	D	E	F	O/D	sing	Max	Mean	P	Sd	RSK	[Lo	Rit	Hi]	Rir	AR	
1	1	C	23	22	42*	9	4		1	5		1	0,42	42	0,49	0,49	20	28	36	25	82	
2	1	E	18	25	25	22	7*		2	12		1	0,07	7	0,25	0,25	0	9	17	7	82	
3	1	A	19*	26	22	17	12		4	19		1	0,19	19	0,40	0,40	-7	0	7	-3	82	
4	1	D	20	27	13	23*	13		4	19		1	0,23	23	0,42	0,42	-5	3	11	0	82	
5	1	B	20	21*	18	17	20		4	20		1	0,21	21	0,41	0,41	-4	5	14	2	82	
6	1	D	20	29	18	18*	13		2	10		1	0,18	18	0,39	0,39	0	9	18	6	82	
7	1	E	16	26	28	17	10*		3	17		1	0,10	10	0,30	0,30	17	27	37	25	82	
8	1	A	24*	24	24	17	7		3	17		1	0,24	24	0,43	0,43	0	8	17	5	82	
9	1	C	24	20	24*	20	8		3	16		1	0,24	24	0,43	0,43	-3	6	14	3	82	
10	1	B	21	22*	24	16	12		5	27		1	0,22	22	0,42	0,42	-14	-5	3	-8	82	

As indicated earlier, the two 187 item tests included two sections of scrambled text (each containing five items). Therefore, each five-item scrambled text section had the potential to be used for the final 60-item test which the panel was seeking to create. Only one section of five suitable items could be selected for each test based on their combined performance (regarding their P and Rit values) and specifications of the anticipated 60-item test. Since the ENG-TALA was used as a model for the Sesotho tests, these specifications are necessarily based on the specifications that were used for the ENG-TALA. The specification for the scrambled text section stated that learners

needed to display knowledge of the relations between the different parts of the text and have the ability to identify the sequence in which different elements of the text follow each other. As indicated in Table 6.10 (for the T-SES-TALA), the two sets of five items each (based on different topics) for the scrambled text section totalled 10 items for scrambled text. The topic of the first five items for the T-SES-TALA was, “*Bakeng sa lerato la di-strawberrie*” (“For the love of strawberries”) and the topic for the other five items, “*MSG – mphato wa dijo kapa motlatselleltsi ya sa utlwisisweng?*” (“MSG – food baddies or misunderstood enhancer?”). As indicated in Table 6.11, the same distribution applied to the NC-SES-TALA. It also contained two sets of five items each (also on different topics) for the scrambled text section totalling 10 items. The topic for the first five items was “*Noka ya Senqu*” (“The Orange River”) and the next five items, “*Mouwane wa metsi o ka fana ka matla a mangata a ntlafaditsweng*” (“Steam can supply clean additional energy”). From the information contained in Tables 6.10 and 6.11, the panel deduced that the first five-item section from the T-SES-TALA, based on the P and the Rit values, performed better than all other five-item scrambled text sections shown in these two tables. An explanation of what constitutes acceptable measures for the P and the Rit values was provided in Section 6.2.3. However, since the test identified for refinement (the NC-SES-TALA) was based on a different theme, according to the principles of TALA development, sections from the two tests could not be used interchangeably. As alluded to above, the theme of the T-SES-TALA was based on food that has both economic and nutritional value, while the theme of the NC-SES-TALA was based on water sources and the value of water power generation. It would, therefore, have been inappropriate to mix test items that had better P and Rit values from the scrambled text with the food theme, with test items from the scrambled text on water sources and power generation. This would have implied violating the principles of the development of the TALA. For this reason, the second best performing scrambled text set which is items 6–10 from the NC-SES-TALA was selected to be used in the final sixty items.

As an additional dimension to the above argument, information in the following convergence tables (Table 6.12 and 6.13) added impetus to the argument that the

second scrambled text section of the NC-SES-TALA should be used for the refined test. Table 6.12 (produced for the T-SES-TALA) shows that from the first Prox iteration of 513 persons, 122 persons were discarded by the program. It produced a person and item reliability of ,17 and ,99 respectively, which shows poor reliability by T-SES-TALA as compared to NC-SES-TALA. Table 6.13 (for the NC-SES-TALA) shows that from the first Prox iteration of 518 persons, 147 persons were discarded to produce a person and item reliability of ,00 and ,98 respectively. This confirms what was deduced from the *TiaPlus* report above that the first scrambled text subtest of the T-SES-TALA performed the best when compared to the other three scrambled texts. However, as stated before, because of the thematic nature of the two tests, the second best scrambled text from the four scrambled text sections (the best performing one from the NC-SES-TALA) had to be selected to be included in the refined NC-SES-TALA.

Table 6.12: T-SES-TALA subtest 1

```
Temporary Workfile Directory: C:\Users\WeidemanAJ\AppData\Local\Temp\
Reading Control Variables ..
Input in process:
Input Data Record:
10011 2
^I ^N^P
513 PERSON Records Input
Building Index ...
      CONVERGENCE TABLE
-Control: Sesotho1_Subtest1_Winsteps.txt Output: \ICELDA\ZOU865WS.TXT
| PROX      ACTIVE COUNT      EXTREME 5 RANGE      MAX LOGIT CHANGE |
| ITERATION PERSON  ITEM CATS      PERSON  ITEM      MEASURES STRUCTURE|
>=====<
| 1    513      5  2      2.77      .83      -1.3863      |
>=====<
| 2    391      5  2      3.02      1.19      -.5378      |
Index Size: 6 with 1 Patterns
Probing data connection: to skip out: Ctrl+F - to bypass: subset=no
Processing unanchored persons ...
>=====<
Data fully connected. No subsets found
-Control: Sesotho1_Subtest1_Winsteps.txt Output: \ICELDA\ZOU865WS.TXT
| JMLE      MAX SCORE      MAX LOGIT      LEAST CONVERGED      CATEGORY STRUCTURE|
| ITERATION RESIDUAL*      CHANGE PERSON  ITEM CAT      RESIDUAL  CHANGE|
>=====<
| 1      -4.70      .5355      6  4*      |
>=====<
| 2      2.60      -.1888      6  1*      |
>=====<
| 3      1.62      -.0727      6  1*      |
>=====<
| 4      .89      -.0306      6  1*      |
>=====<
| 5      .48      -.0139      9  1*      |
>=====<
| 6      .26      -.0067      9  1*      |
>=====<
| 7      .14      -.0034      9  1*      |
```

```

-----
Calculating JMLE Fit Statistics
>=====<
Time for estimation: 0:0:1.499
Output to C:\Albert Documents\ICELDA\ZOU865WS.TXT
ice02716_TALA Sesotho1_Subtest1_Winsteps.xlsx
-----
| PERSON   513 INPUT   513 MEASURED      INFIT      OUTFIT |
|   TOTAL   COUNT   MEASURE REALSE   IMNSQ   ZSTD OMNSQ   ZSTD|
| MEAN    1.6    5.0   -1.04  1.40    1.00   .0  1.02   .1|
| P.SD    1.3    .0    1.59  .37    .50   .9  .90   .8|
| REAL RMSE 1.45 TRUE SD  .65 SEPARATION .45 PERSON RELIABILITY .17|
|-----|
| ITEM     5 INPUT     5 MEASURED      INFIT      OUTFIT |
|   TOTAL   COUNT   MEASURE REALSE   IMNSQ   ZSTD OMNSQ   ZSTD|
| MEAN   167.4   513.0    .00  .13    1.00  -.2  1.02   .0|
| P.SD   76.2    .0    1.09  .02    .09  1.7  .13  1.7|
| REAL RMSE .13 TRUE SD  1.09 SEPARATION 8.32 ITEM RELIABILITY .99|
|-----|

Output written to C:\Albert Documents\ICELDA\ZOU865WS.TXT
CODES= 01
Measures constructed: use "Diagnosis" and "Output Tables" menus

Extra specifications (if any). Press Enter to analyze:

```

Table 6.13: NC-SES-TALA subtest 1

```

Temporary Workfile Directory: C:\Users\WeidemanAJ\AppData\Local\Temp\
Reading Control Variables ..
Input in process:
Input Data Record:
00000 2
^I ^N^P
518 PERSON Records Input
Building Index ...
      CONVERGENCE TABLE
-Control: Sesotho2_Subtest1_Winsteps.txt Output: \ICELDA\ZOU348WS.TXT
| PROX      ACTIVE COUNT      EXTREME 5 RANGE      MAX LOGIT CHANGE |
| ITERATION PERSON  ITEM CATS    PERSON  ITEM    MEASURES STRUCTURE|
|-----|
| 1    518    5    2    2.45    .84    -1.3863    |
|-----|
| 2    371    5    2    2.68    1.04    -.3108    |
|-----|
Index Size: 5 with 1 Patterns
Probing data connection: to skip out: Ctrl+F - to bypass: subset=no
Processing unanchored persons ...
>=====<
Data fully connected. No subsets found
-Control: Sesotho2_Subtest1_Winsteps.txt Output: \ICELDA\ZOU348WS.TXT
| JMLE    MAX SCORE    MAX LOGIT    LEAST CONVERGED    CATEGORY STRUCTURE|
| ITERATION RESIDUAL*    CHANGE PERSON  ITEM CAT  RESIDUAL  CHANGE|
|-----|
| 1    -4.61    -.4285    1    3*    |
|-----|
| 2    -2.00    -.1476    1    3*    |
|-----|
| 3    -.89    -.0511    1    3*    |
|-----|
| 4    -.41    -.0186    1    3*    |
|-----|
| 5    -.19    -.0072    1    3*    |
|-----|
| 6    -.09    -.0029    3    3*    |
|-----|

Calculating JMLE Fit Statistics
>=====<
Time for estimation: 0:0:1.450
Output to C:\Albert Documents\ICELDA\ZOU348WS.TXT
ice02686_Sesotho2_Subtest1_Winsteps.xlsx

```

```

-----
| PERSON  518 INPUT    518 MEASURED      INFIT      OUTFIT |
|      TOTAL  COUNT  MEASURE REALSE    IMNSQ  ZSTD OMNSQ  ZSTD|
| MEAN    1.1    5.0   -1.59  1.41    1.00   .1   .99   .1|
| P.SD    1.0    .0    1.17   .35     .41   .8   .79   .7|
| REAL RMSE 1.45 TRUE SD   .00 SEPARATION .00 PERSON RELIABILITY .00|
|-----|
| ITEM    5 INPUT    5 MEASURED      INFIT      OUTFIT |
|      TOTAL  COUNT  MEASURE REALSE    IMNSQ  ZSTD OMNSQ  ZSTD|
| MEAN   117.2   518.0    .00   .13    1.00   .1   .99  -.1|
| P.SD    59.2    .0     .94   .02     .03   .6   .04   .5|
| REAL RMSE .14 TRUE SD   .93 SEPARATION 6.89 ITEM RELIABILITY .98|
|-----|

Output written to C:\Albert Documents\ICELDA\ZOU348WS.TXT
CODES= 01
Measures constructed: use "Diagnosis" and "Output Tables" menus

```

As seen above in Tables 6.8 and 6.9, the NC-SES-TALA overall produced better performing subtests than the T-SES-TALA (besides scrambled texts). For this reason, the number of items for the other subtests was reduced following a similar procedure to that employed for the scrambled text section (as displayed in Tables 6.10 and 6.11). The first subtest to be reduced after scrambled text, was subtest number three (text comprehension), by utilising the data shown in Table 6.14.

Table 6.14: NC-SES-TALA *TiaPlus* report for items making up subtest 3

Persons: All persons Subtest(3): Text comprehension
11-83

Test and Item Analysis

		% #																			
		----- P- and A- values -----										Mis- ----- Weighted -----									
Item	Item																				
Label	nr.	Weight	Key	A	B	C	D	E	F	O/D	sing	Max	Mean	P	Sd	RSK	[Lo	Rit	Hi]	Rir	AR

1	11	1	C	23	23	41*	10	1		1	6	1	0,41	41	0,49	0,49	32	39	47	35	86
2	12	1	D	14	20	25	38*	2		1	7	1	0,38	38	0,49	0,49	35	42	49	38	86
3	13	1	A	24*	29	26	20	1		1	7	1	0,24	24	0,42	0,42	4	14	23	10	86
4	14	1	A	32*	25	24	15	2		2	10	1	0,32	32	0,47	0,47	27	36	44	32	86
5	15	1	C	16	18	44*	17	3		3	14	1	0,44	44	0,50	0,50	45	51	56	47	86
6	16	1	C	20	16	45*	17	1		1	4	1	0,45	45	0,50	0,50	43	49	54	45	86
7	17	1	D	21	21	22	34*	0		2	10	1	0,34	34	0,47	0,47	45	52	58	48	86
8	18	1	A	26*	25	33	12	1		3	13	1	0,26	26	0,44	0,44	10	19	28	15	86
9	19	1	D	19	21	20	36*	2		2	11	1	0,36	36	0,48	0,48	40	47	54	43	86
10	20	1	C	20	21	38*	14	3		3	17	1	0,38	38	0,49	0,49	44	50	56	46	86
11	21	1	C	37	23	24*	12	2		2	12	1	0,24	24	0,43	0,43	13	22	31	18	86
12	22	1	B	15	43*	27	12	1		2	12	1	0,43	43	0,50	0,50	35	42	49	38	86
13	23	1	D	19	31	19	26*	2		3	13	1	0,26	26	0,44	0,44	19	28	37	24	86
14	24	1	D	19	25	24	29*	2		2	11	1	0,29	29	0,45	0,45	32	41	48	37	86
15	25	1	C	18	21	39*	16	2		3	18	1	0,39	39	0,49	0,49	38	45	52	41	86
16	26	1	C	27	22	35*	14	1		1	7	1	0,35	35	0,48	0,48	55	61	66	58	86
17	27	1	B	30	31*	25	9	1		3	14	1	0,31	31	0,46	0,46	-6	2	10	-2	87
18	28	1	C	16	23	45*	13	2		2	8	1	0,45	45	0,50	0,50	44	49	55	46	86
19	29	1	C	20	23	34*	18	2		2	12	1	0,34	34	0,48	0,48	40	47	54	43	86
20	30	1	C	26	22	32*	15	2		3	13	1	0,32	32	0,47	0,47	26	34	42	30	86
21	31	1	D	28	22	20	26*	2		2	8	1	0,26	26	0,44	0,44	34	43	50	39	86

22	32	1	B	13	44*	27	13	0	2 11 1 0,44	44	0,50	0,50	34	41	47	37	86
23	33	1	C	19	31	30*	17	0	2 12 1 0,30	30	0,46	0,46	7	15	24	11	86
24	34	1	C	17	22	40*	16	1	3 16 1 0,40	40	0,49	0,49	31	38	45	34	86
25	35	1	B	26	38*	18	13	3	3 14 1 0,38	38	0,49	0,49	37	46	51	40	86
26	36	1	D	35	19	17	27*	1	2 8 1 0,27	27	0,44	0,44	25	34	42	30	86
27	37	1	C	28	29	27*	15	0	2 8 1 0,27	27	0,44	0,44	5	14	23	10	86
28	38	1	C	15	29	42*	12	1	2 11 1 0,42	42	0,49	0,49	27	35	42	31	86
29	39	1	C	25	24	32*	16	1	3 14 1 0,32	32	0,47	0,47	15	24	32	19	86
30	40	1	C	27	21	29*	18	3	3 15 1 0,29	29	0,45	0,45	20	29	38	25	86
31	41	1	C	30	19	35*	13	1	2 8 1 0,35	35	0,48	0,48	36	45	50	39	86
32	42	1	B	20	40*	26	11	1	2 8 1 0,40	40	0,49	0,49	8	16	25	12	86
33	43	1	B	25	37*	21	14	1	2 12 1 0,37	37	0,48	0,48	17	25	33	21	86
34	44	1	B	18	33*	26	18	1	4 21 1 0,33	33	0,47	0,47	16	25	34	21	86
35	45	1	C	18	24	36*	16	2	4 22 1 0,36	36	0,48	0,48	38	46	52	41	86
36	46	1	C	27	24	33*	12	1	3 14 1 0,33	33	0,47	0,47	34	46	49	38	86
37	47	1	B	18	34*	30	14	1	2 12 1 0,34	34	0,48	0,48	8	16	25	12	86
38	48	1	C	18	30	34*	13	2	3 14 1 0,34	34	0,47	0,47	5	14	22	9	86
39	49	1	D	20	33	21	23*	1	3 16 1 0,23	23	0,42	0,42	22	32	41	28	86
40	50	1	C	24	21	31*	18	2	3 18 1 0,31	31	0,46	0,46	32	40	47	36	86
41	51	1	A	46*	22	16	12	2	3 14 1 0,46	46	0,50	0,50	20	28	35	23	86
42	52	1	D	10	31	28	27*	1	3 15 1 0,27	27	0,44	0,44	43	51	58	47	86
43	53	1	C	15	30	36*	16	1	3 16 1 0,36	36	0,48	0,48	15	24	32	19	86
44	54	1	C	14	25	33*	23	1	3 17 1 0,33	33	0,47	0,47	29	37	45	33	86
45	55	1	B	22	35*	20	17	2	4 20 1 0,35	35	0,48	0,48	35	46	50	39	86
46	56	1	A	39*	25	18	13	2	3 16 1 0,39	39	0,49	0,49	28	35	43	31	86
47	57	1	D	15	34	27	20*	2	2 11 1 0,20	20	0,40	0,40	0	9	19	5	86
48	58	1	A	26*	26	32	12	1	3 18 1 0,26	26	0,44	0,44	24	33	42	29	86
49	59	1	A	27*	28	22	18	2	3 16 1 0,27	27	0,45	0,45	24	33	41	29	86
50	60	1	D	21	26	22	25*	1	4 23 1 0,25	25	0,43	0,43	21	30	39	26	86
51	61	1	A	37*	25	20	14	1	3 14 1 0,37	37	0,48	0,48	7	16	25	12	86
52	62	1	C	14	34	31*	15	1	4 23 1 0,31	31	0,46	0,46	1	10	19	5	86
53	63	1	A	17*	33	31	15	1	3 15 1 0,17	17	0,38	0,38	-3	5	14	2	86
54	64	1	A	21*	29	23	20	2	4 22 1 0,21	21	0,41	0,41	0	9	18	5	86
55	65	1	C	24	21	30*	18	1	6 29 1 0,30	30	0,46	0,46	21	30	38	26	86
56	66	1	B	28	29*	23	15	1	4 20 1 0,29	29	0,45	0,45	16	25	34	21	86
57	67	1	C	13	28	39*	16	0	3 16 1 0,39	39	0,49	0,49	31	38	45	34	86
58	68	1	B	21	32*	28	15	1	3 18 1 0,32	32	0,46	0,46	3	12	21	8	86
59	69	1	D	18	32	27	19*	1	3 14 1 0,19	19	0,39	0,39	-9	-1	6	-5	87
60	70	1	B	25	31*	23	16	2	4 19 1 0,31	31	0,46	0,46	12	20	29	16	86
61	71	1	B	30	25*	21	18	1	5 24 1 0,25	25	0,44	0,44	-2	6	15	2	87
62	72	1	C	13	31	37*	13	1	5 24 1 0,37	37	0,48	0,48	24	32	40	28	86
63	73	1	A	24*	30	26	15	1	4 23 1 0,24	24	0,43	0,43	17	26	35	22	86
64	74	1	A	22*	24	30	19	1	4 23 1 0,22	22	0,41	0,41	16	25	34	21	86
65	75	1	D	27	23	19	24*	2	5 28 1 0,24	24	0,43	0,43	30	40	48	36	86
66	76	1	C	34	23	26*	13	1	3 13 1 0,26	26	0,44	0,44	29	38	46	34	86
67	77	1	B	13	39*	27	15	1	4 23 1 0,39	39	0,49	0,49	14	22	31	18	86
68	78	1	B	17	37*	29	13	0	4 19 1 0,37	37	0,48	0,48	8	16	25	12	86
69	79	1	D	21	26	24	25*	1	3 14 1 0,25	25	0,43	0,43	10	19	28	15	86
70	80	1	B	30	28*	24	12	2	4 22 1 0,28	28	0,45	0,45	5	14	22	9	86
71	81	1	C	31	29	21*	14	1	4 20 1 0,21	21	0,41	0,41	3	13	23	9	86
72	82	1	A	31*	25	29	10	1	4 21 1 0,31	31	0,46	0,46	36	46	51	40	86
73	83	1	B	25	35*	24	12	1	3 17 1 0,35	35	0,48	0,48	14	23	31	18	86

SubGroup number	: 0	SubTest number	: 3
Number of persons in test	: 519	Number of selected items	: 73
Minimum test score	: 0	Maximum test score	: 73
Average test score	: 23,45	Standard deviation	: 10,25
Average P-value	: 32,12	Std. Error of Measurement	: 3,79
Average Rit	: 0,31		
Coefficient Alpha	: 0,86	SE Coeff. Alpha	: 0,01
GLB	: 0,94	Asymptotic GLB coef	: Unavailable
Items used in GLB proc	: 73		

 90% Confidence limits for Coefficient Alpha: (0,85 =< 0,86 =< 0,88)

Estimated Coefficient Alpha if this test had a standard norm length of 40 items: 0,78 (Spearman-Brown)

Table 6.14 includes all items from subtest 3. The initial test contained 73 items from which 20 had to be selected. Again, these items were selected based on their specifications and their combined performance. These specifications (see Table 6.15), in combination with the P and the Rit values, indicate the validity of the construct of the test. This is an important point to make since it emphasises that these specifications contribute to the overall validity of the test. The specifications for the items in subtest 3 were evaluated against the specifications that were used for the ENG-TALA. Based on this evaluation, the following table of specifications (Table 6.15) was generated for a 60-item NC-SES-TALA test.

Table 6.15: Specifications for subtest 3

Subtest	Specifications for 60-item test including the number of items selected for each measured ability
Text Comprehension	The 20 items to be included in the refined test: √ inferencing /extrapolation (3) √ distinction (4) √ sequencing (3) √ comparing text with text (3) √ handling metaphor (1) √ communicative function (1) √ vocabulary in context (1) √ categorisation (1) √ cohesion (3)

The highlighted items (in yellow) in Table 6.14 show better productivity than the unmarked items in terms of P and Rit values (marked in green). As explained before, the P value represents the proportion of learners who answered the specific item correctly (again, for a test such as this, the P value should be above 0,15 and below 0,84). With regard to the Rit value, a high Rit means that learners with high ability tend to answer the item correctly and that learners with a low ability tend to answer the item incorrectly – the item in other words discriminates well. When this happens, the item is said to be consistent with the test. With a Rit of zero, the implication is that there is no relation whatsoever between the performance of the test as a whole and the performance of the specific test item. However, with a negative Rit, learners with a high ability tend to answer the item incorrectly, but learners with a low ability tend to answer the item correctly. According to Cito (2013:36), a high Rit value indicates that the item contributes substantially to the test's reliability and *TiaPlus* displays the Rit values as multiplied by 100 (Cito, 2013:36). So, based on the values shown in Table 6.14, as well as specifications that meet the criteria for the construct for this kind of a test, a number of items which are viewed as consistent with the test have been marked for possible inclusion into the final 60- item test. It is also important at this point to explain the value in the AR column at the extreme right of the table. AR stands for Alpha Rest coefficient and it is a value of how the alpha coefficient of the test is affected when a specific test item is removed (Cito, 2013:30). The AR value therefore provides information on how the removal of specific items would affect the overall reliability of the subtest in the test. One would therefore not wish to remove items that would lower the reliability of the test. As can be seen in Table 6.14, the AR value of all items in subtest 3 is 86 or higher, which means that it is equivalent to or slightly higher than the average coefficient alpha of the whole subtest (0,86). The AR value should preferably not be lower than the coefficient alpha of the test. According to Cito (2013:30), removing an item with an AR that is lower than the coefficient alpha of the test, has the potential to render the test with remaining items less reliable. The panel therefore also had to take note of this attribute while reducing the items to sixty. It was only for the scrambled text section where this aspect could not be controlled as the whole set of five items as a unit had to be identified for inclusion into the final test.

However, as already mentioned, only 20 items from the comprehension section (cf. Table 6.15) could be considered for inclusion in the final 60-item test. Additionally, the following *Winsteps* analysis (Table 6.16) shows that only one participant's results were excluded from this subtest. According to Green (2013:152), participants' results are excluded from analyses on condition that they get all items correct or they get all items incorrect. The reason for this is that the *Winsteps* model cannot accurately measure such a person's ability level. By excluding such participants, it makes the subtest fit for the population for which it was used and the person reliability as well as the item reliability appropriate at ,83 and ,92 respectively.

Table 6.16: NC-SES-TALA subtest 3

```
518 PERSON Records Input
Building Index ...
      CONVERGENCE TABLE
-Control: Sesotho2_Subtest3_Winsteps.txt Output: \ICELDA\ZOU738WS.TXT
| PROX      ACTIVE COUNT      EXTREME 5 RANGE      MAX LOGIT CHANGE |
| ITERATION PERSON  ITEM CATS      PERSON  ITEM      MEASURES STRUCTURE|
>=====<
|   1   518   73   2   3.70  1.20      -4.2767      |
>=====<
|   2   517   73   2   3.77  1.29      -.0844      |
Index Size: 51 with 1 Patterns
Probing data connection: to skip out: Ctrl+F - to bypass: subset=no
Processing unanchored persons ...
>=====<
Data fully connected. No subsets found
|Control: Sesotho2_Subtest3_Winsteps.txt Output: \ICELDA\ZOU738WS.TXT
| JMLE      MAX SCORE MAX LOGIT      LEAST CONVERGED CATEGORY STRUCTURE|
| ITERATION RESIDUAL*  CHANGE PERSON  ITEM CAT  RESIDUAL  CHANGE|
>=====<
|   1      -1.52   -.3302   237   59*      |
>=====<
|   2      -.58   -.1003    9   59*      |
>=====<
|   3      -.21   -.0305    9   53*      |
>=====<
|   4      -.08   -.0093   10   53*      |
-----
Calculating JMLE Fit Statistics
>=====<
Time for estimation: 0:0:1.367
Output to C:\Albert Documents\ICELDA\ZOU738WS.TXT
ice02694_Sesotho2_Subtest3_Winsteps.xlsx
-----
| PERSON   518 INPUT   518 MEASURED      INFIT      OUTFIT |
|   TOTAL   COUNT   MEASURE REALSE      IMNSQ  ZSTD OMNSQ  ZSTD|
| MEAN   23.4   73.0   -.83   .28   1.00  -.1  1.01   .0|
| P.SD   10.3    .0    .70   .08   .05   .6  .07   .7|
| REAL RMSE .29 TRUE SD  .64 SEPARATION 2.23 PERSON RELIABILITY .83|
|-----|
```

ITEM	73 INPUT	73 MEASURED	INFIT		OUTFIT		
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD OMNSQ	ZSTD
MEAN	166.4	518.0	.00	.10	1.00	-.3	1.01
P.SD	35.7	.0	.36	.01	.09	2.3	.11
REAL RMSE	.10	TRUE SD	.34	SEPARATION	3.34	ITEM RELIABILITY .92	

Output written to C:\Albert Documents\ICELDA\ZOU738WS.TXT
CODES= 01
Measures constructed: use "Diagnosis" and "Output Tables" menus

Following the above-mentioned reduction of the number of items for subtest 3 (text comprehension), the panel also engaged into reducing the items for subtest four (interpreting graphs and visual information). The same procedure as above was followed for this subtest. In Table 6.18, items belonging to subtest four (interpreting graphs and visual information) are displayed. Similar to Table 6.14, items highlighted in yellow are the items which displayed better productivity in both the *TiaPlus* and *Winsteps* analyses. These items were, therefore, selected for inclusion into the sixty item test. Item number 30 below in Table 6.18 could not be excluded because of its AR value. As explained before, the Alpha-Rest (AR) value for an item refers to the Cronbach's Alpha Coefficient of overall test (or section, in this case) reliability calculated for the test/section with the item removed (Cito, 2013:36). In the decision about which eight items to include in this subtest, Items 27 or 28 would have been a better choice to include, was it not for the AR value of item 30. The alpha of items 27 and 28 are 32 for both, while the alpha of item 30 is 31. This implies that if item 30 is removed, the alpha of the subtest is likely to go down from 33 to 31. This is confirmation that it was not wise to remove item 33 at the expense of both items 27 and 28. The productivity of selected items for subtest four was based on their specifications and combined performance in terms of their P and Rit values. Their specifications are captured in Table 6.17 for a 60-item NC-SES-TALA.

Table 6.17: Specifications for subtest 4

Subtest	Specifications for 60-item test including number of items selected for each measured ability
Interpreting graphs and visual information	(8) √ compare (3) √ proportions (1) √ trends (2) √ averages (1) √ comparison between combined readings (1)

Table 6.18: NC-SES-TALA TiaPlus report for items making up subtest 4

Persons: All persons Subtest(4): Interpreting graphs & visual info
84-113

Test and Item Analysis

		% #																							
		----- P- and A- values -----										Mis- ----- Weighted -----													
Item	Item	nr.	Weight	Key	A	B	C	D	E	F	O/D	sing	Max	Mean	P	Sd	RSK	[Lo	Rit	Hi]	Rir	AR			

1	84	1	B	27	24*	32	14	0					2	11		1	0,24	24	0,43	0,43	20	28	36	13	31
2	85	1	A	10*	38	32	15	0					4	22		1	0,10	10	0,30	0,30	-1	9	19	-1	34
3	86	1	B	11	24*	45	16	0					4	19		1	0,24	24	0,43	0,43	6	15	23	0	34
4	87	1	C	14	24	33*	24	0					5	28		1	0,33	33	0,47	0,47	17	25	33	8	32
5	88	1	B	18	24*	29	20	0					9	48		1	0,24	24	0,43	0,43	17	25	33	11	32
6	89	1	D	22	24	34	15*	0					4	23		1	0,15	15	0,36	0,36	9	18	27	5	33
7	90	1	B	11	31*	37	18	0					3	18		1	0,31	31	0,46	0,46	13	21	30	5	33
8	91	1	C	14	28	38*	15	0					5	28		1	0,38	38	0,49	0,49	17	25	33	9	32
9	92	1	C	14	27	33*	21	0					5	28		1	0,33	33	0,47	0,47	21	29	36	13	31
10	93	1	B	17	30*	29	17	0					7	36		1	0,30	30	0,46	0,46	30	37	45	22	29
11	94	1	A	25*	30	27	13	0					4	19		1	0,25	25	0,43	0,43	13	22	30	7	32
12	95	1	B	11	33*	40	15	0					2	12		1	0,33	33	0,47	0,47	13	21	29	5	33
13	96	1	A	15*	33	30	18	0					3	15		1	0,15	15	0,36	0,36	7	16	25	4	33
14	97	1	C	15	26	33*	23	0					4	20		1	0,33	33	0,47	0,47	24	32	39	16	30
15	98	1	C	16	26	33*	17	0					9	45		1	0,33	33	0,47	0,47	9	18	26	1	34
16	99	1	B	26	24*	29	16	0					5	28		1	0,24	24	0,43	0,43	10	19	27	4	33
17	100	1	A	11*	37	33	15	0					5	24		1	0,11	11	0,31	0,31	4	14	23	3	33
18	101	1	D	12	28	39	16*	0					5	25		1	0,16	16	0,37	0,37	1	10	19	-3	34

19	102	1	C	15	31	30*	18	0	6	31		1	0,30	30	0,46	0,46	20	28	36	12	31
20	103	1	C	17	21	35*	20	0	7	38		1	0,35	35	0,48	0,48	17	25	33	9	32
21	104	1	C	21	17	34*	21	0	7	36		1	0,34	34	0,47	0,47	11	19	27	2	34
22	105	1	C	11	32	36*	16	0	5	26		1	0,36	36	0,48	0,48	12	20	28	3	33
23	106	1	D	12	31	35	17*	0	5	27		1	0,17	17	0,37	0,37	15	24	33	11	32
24	107	1	A	16*	28	31	19	0	6	31		1	0,16	16	0,37	0,37	10	20	29	7	32
25	108	1	A	17*	22	32	21	0	8	39		1	0,17	17	0,38	0,38	6	15	25	2	33
26	109	1	B	23	25*	31	16	0	5	25		1	0,25	25	0,43	0,43	10	19	27	4	33
27	110	1	B	9	37*	36	13	0	5	25		1	0,37	37	0,48	0,48	17	24	32	8	32
28	111	1	D	14	32	31	17*	0	5	25		1	0,17	17	0,38	0,38	15	24	33	11	32
29	112	1	A	16*	27	34	18	0	5	28		1	0,16	16	0,36	0,36	15	24	33	12	32
30	113	1	C	18	29	27*	21	0	6	30		1	0,27	27	0,44	0,44	19	27	35	12	31

```

-----
SubGroup number      : 0          SubTest number      : 4
Number of persons in test : 519      Number of selected items : 30
Minimum test score      : 0          Maximum test score      : 30
Average test score      : 7,59      Standard deviation      : 2,83
Average P-value        : 25,29      Std. Error of Measurement : 2,32
Average Rit            : 0,22
Coefficient Alpha        : 0,33      SE Coeff. Alpha        : 0,04
GLB                     : 0,59      Asymptotic GLB coef    : Unavailable
Items used in GLB proc  : 30

```

```

-----
90% Confidence limits for Coefficient Alpha: (0,26 =< 0,33 =< 0,40)
-----

```

```

-----
Estimated Coefficient Alpha if this test had a standard norm length of 40 items: 0,40 (Spearman-Brown)
-----

```

With regard to the *Winsteps* analysis for subtest 4, no major discrepancies are seen for item reliability in the *Winsteps* convergence table below (Table 6.19). Only two persons' results were excluded from the analysis (reducing the number of test takers from 518 to 516), something showing that the population was also fit for the subtest. Moreover, the program had undergone only two iterations (quick estimates) in order to produce the results, which is an indication that the items perform well in terms of the mathematical algorithms used for the analyses – the more iterations necessary, the worse the items. The infit and outfit measures are also within the required parameters of 0,5 and 1,5 (cf. Section 6.2.3 for the definition of infit and outfit measures). The only limitation shown by this table is person reliability, which stands at a low ,36 as compared with subtests 3, 5, and 6. The person reliability of these three subtests stands at ,83, ,64, and ,63, respectively. With these high person reliabilities, the impression is created that these subtests could be replicated many times more than subtest 4 and produce the same results, if the same test-takers were used.

Table 6.19: NC-SES-TALA subtest 4

```

518 PERSON Records Input
Building Index ...
      CONVERGENCE TABLE
-Control: Sesotho2_Subtest4_Winsteps.txt Output: \ICELDA\ZOU014WS.TXT
|  PROX      ACTIVE COUNT      EXTREME 5 RANGE      MAX LOGIT CHANGE |
|  ITERATION PERSON  ITEM CATS      PERSON  ITEM      MEASURES STRUCTURE|
>=====<
|    1    518    30    2    3.54  1.32    -3.3673      |
>=====<
|    2    516    30    2    3.70  1.40    -1.1450      |
Index Size: 19 with 1 Patterns
Probing data connection: to skip out: Ctrl+F - to bypass: subset=no
Processing unanchored persons ...
>=====<
Data fully connected. No subsets found
|Control: Sesotho2_Subtest4_Winsteps.txt Output: \ICELDA\ZOU014WS.TXT
|  JMLE      MAX SCORE      MAX LOGIT      LEAST CONVERGED      CATEGORY STRUCTURE|
|  ITERATION RESIDUAL*      CHANGE PERSON  ITEM  CAT  RESIDUAL  CHANGE|
>=====<
|    1      -1.70    -0.3141    141    15*      |
>=====<
|    2      -0.49    -0.0941    141     7*      |
>=====<
|    3      -0.13    -0.0281    141     7*      |
>=====<
|    4      -0.04    -0.0084    141     7*      |
-----
Calculating JMLE Fit Statistics
>=====<
Time for estimation: 0:0:1.341
Output to C:\Albert Documents\ICELDA\ZOU014WS.TXT
ice02698_Sesotho2_Subtest4_Winsteps.xlsx
-----
|  PERSON    518 INPUT    518 MEASURED      INFIT      OUTFIT  |
|  TOTAL    COUNT    MEASURE REALSE      IMNSQ  ZSTD OMNSQ  ZSTD|
|  MEAN     7.6    30.0    -1.22   .46    1.00   .1  1.01   .1|
|  P.SD     2.8     .0     .60   .12     .08   .5  .20   .7|
|  REAL RMSE .48 TRUE SD   .36 SEPARATION .75 PERSON RELIABILITY .36|
|-----|
|  ITEM     30 INPUT     30 MEASURED      INFIT      OUTFIT  |
|  TOTAL    COUNT    MEASURE REALSE      IMNSQ  ZSTD OMNSQ  ZSTD|
|  MEAN    131.0   518.0     .00   .11    1.00   .0  1.01   .0|
|  P.SD    43.0     .0     .51   .01     .02   .5  .05   .7|
|  REAL RMSE .11 TRUE SD   .49 SEPARATION 4.50 ITEM RELIABILITY .95|
|-----|
Output written to C:\Albert Documents\ICELDA\ZOU014WS.TXT
CODES= 01
Measures constructed: use "Diagnosis" and "Output Tables" menus

```

Another subtest that needed a reduction of items was subtest 5 (Table 6.20) which focuses on vocabulary. The subtest had a total of 32 items which needed to be 165

reduced to only 15 items. Items that were retained are marked in yellow as indicated in Table 6.21. Items were evaluated based on the same principles that were used for those in Tables 6.14 and 6.18. A *Winsteps* convergence table is also included to provide further support for the judgement about the selection of specific items for this subtest. The item specifications for a 60-item NC-SES-TALA subtest 5 are provided below.

Table 6.20: Specifications for subtest 5

Subtest	Specifications for 60-item test including number of items selected for each measured ability
Vocabulary knowledge	(15) √ vocabulary in context (8) √ Handling idiom (7)

Table 6.21: NC-SES-TALA *TiaPlus* report for items making up subtest 5

Persons: All persons

Subtest(5): Vocabulary knowledge

114-145

Test and Item Analysis

		% #																						
		----- P- and A- values -----										Mis- ----- Weighted -----												
Item	Item	nr.	Weight	Key	A	B	C	D	E	F	O/D	sing	Max	Mean	P	Sd	RSK	[Lo	Rit	Hi]	Rir	AR		

1	114	1	B	21	30*	31	14	0			5	25		1	0,30	30	0,46	0,46	29	37	44	27	66	
2	115	1	D	9	32	33	22*	0			4	23		1	0,22	22	0,41	0,41	34	42	50	34	66	
3	116	1	B	13	32*	35	15	0			5	27		1	0,32	32	0,46	0,46	26	33	40	23	66	
4	117	1	A	19*	25	34	16	0			7	34		1	0,19	19	0,39	0,39	33	42	50	34	66 x	
5	118	1	C	16	29	30*	18	0			7	37		1	0,30	30	0,46	0,46	4	12	20	1	68	
6	119	1	D	20	23	28	24*	0			6	30		1	0,24	24	0,43	0,43	37	44	52	36	65 x	
7	120	1	B	9	35*	36	14	0			6	32		1	0,35	35	0,48	0,48	17	25	33	14	67	
8	121	1	C	12	28	40*	15	0			6	30		1	0,40	40	0,49	0,49	28	34	41	24	66	
9	122	1	C	11	27	36*	19	0			6	30		1	0,36	36	0,48	0,48	27	34	41	23	66	
10	123	1	C	20	23	33*	18	0			7	35		1	0,33	33	0,47	0,47	22	31	38	20	67	
11	124	1	D	20	23	27	24*	0			5	28		1	0,24	24	0,43	0,43	35	43	50	34	66 x	
12	125	1	B	12	39*	33	12	0			5	26		1	0,39	39	0,49	0,49	27	34	41	23	66	
13	126	1	C	15	29	36*	15	0			5	25		1	0,36	36	0,48	0,48	25	32	39	22	66	
14	127	1	D	13	25	30	25*	0			7	35		1	0,25	25	0,43	0,43	33	40	48	31	66	
15	128	1	B	18	28*	30	18	0			7	37		1	0,28	28	0,45	0,45	30	38	45	28	66	
16	129	1	B	16	32*	31	14	0			7	36		1	0,32	32	0,47	0,47	29	37	43	26	66	
17	130	1	D	14	30	32	18*	0			6	32		1	0,18	18	0,39	0,39	11	20	29	11	67	
18	131	1	C	13	27	36*	17	0			7	36		1	0,36	36	0,48	0,48	35	42	48	31	66	
19	132	1	D	13	25	31	24*	0			7	35		1	0,24	24	0,43	0,43	29	37	45	28	66	
20	133	1	D	22	23	30	17*	0			8	42		1	0,17	17	0,38	0,38	13	23	32	14	67	
21	134	1	D	24	24	26	19*	0			7	35		1	0,19	19	0,39	0,39	31	39	47	31	66	
22	135	1	B	10	34*	34	16	0			6	33		1	0,34	34	0,47	0,47	14	23	31	12	67	
23	136	1	D	16	33	30	15*	0			6	30		1	0,15	15	0,36	0,36	11	20	29	12	67	
24	137	1	C	12	29	35*	18	0			6	31		1	0,35	35	0,48	0,48	4	12	20	1	68	
25	138	1	D	19	23	26	23*	0			8	44		1	0,23	23	0,42	0,42	30	39	47	30	66	
26	139	1	C	20	22	33*	16	0			8	44		1	0,33	33	0,47	0,47	24	32	39	21	66	
27	140	1	D	10	32	34	18*	0			7	35		1	0,18	18	0,38	0,38	19	28	37	20	67	
28	141	1	A	16*	29	32	18	0			6	30		1	0,16	16	0,36	0,36	20	31	40	22	66	
29	142	1	C	14	27	36*	17	0			6	33		1	0,36	36	0,48	0,48	14	22	30	11	67	
30	143	1	D	21	26	28	17*	0			8	40		1	0,17	17	0,38	0,38	1	8	15	-1	68	
31	144	1	B	21	29*	26	18	0			7	36		1	0,29	29	0,45	0,45	18	27	35	16	67	
32	145	1	B	11	30*	31	21	0			8	39		1	0,30	30	0,46	0,46	-5	2	10	-8	69	

SubGroup number				: 0				SubTest number				: 5												
Number of persons in test				: 519				Number of selected items				: 32												
Minimum test score				: 0				Maximum test score				: 32												
Average test score				: 8,85				Standard deviation				: 4,23												
Average P-value				: 27,65				Std. Error of Measurement				: 2,42												
Average Rit				: 0,30																				
Coefficient Alpha				: 0,67				SE Coeff. Alpha				: 0,02												
GLB				: 0,80				Asymptotic GLB coef				: Unavailable												

Items used in GLB proc : 32

90% Confidence limits for Coefficient Alpha: (0,64 =< 0,67 =< 0,71)

Estimated Coefficient Alpha if this test had a standard norm length of 40 items: 0,72 (Spearman-Brown)

In the *Winsteps* convergence table (Table 6.22), it can be seen that the item reliability for this subtest is acceptable at ,94 and person reliability is lower than the required parameter at 6.4. The person reliability was brought down by 8 persons that were removed by the model. However, after 2 iterations, the program was able to produce infit and outfit measures that were well within the required parameters. They were both at 1,00 which indicated that items and persons used in this analysis were fit for this subtest.

Table 6.22: NC-SES-TALA subtest 5

518 PERSON Records Input

Building Index ...

CONVERGENCE TABLE

-Control: Sesotho2_Subtest5_Winsteps.txt Output: \ICELDA\ZOU825WS.TXT

	PROX	ACTIVE	COUNT	EXTREME	5	RANGE	MAX	LOGIT	CHANGE	
	ITERATION	PERSON	ITEM	CATS	PERSON	ITEM	MEASURES	STRUCTURE		

>=====<

	1	518	32	2	4.36	1.11	-3.4340	
--	---	-----	----	---	------	------	---------	--

>=====<

	2	510	32	2	4.49	1.21	-.1000	
--	---	-----	----	---	------	------	--------	--

Index Size: 27 with 1 Patterns

Probing data connection: to skip out: Ctrl+F - to bypass: subset=no

Processing unanchored persons ...

>=====<

Data fully connected. No subsets found

|Control: Sesotho2_Subtest5_Winsteps.txt Output: \ICELDA\ZOU825WS.TXT

JMLE	MAX SCORE	MAX LOGIT	LEAST CONVERGED	CATEGORY STRUCTURE
ITERATION	RESIDUAL*	CHANGE PERSON	ITEM CAT	RESIDUAL CHANGE

>=====<

1	-1.27	-.3323	58	2*
---	-------	--------	----	----

>=====<

2	-.47	-.1008	58	4*
---	------	--------	----	----

>=====<

3	-.17	-.0306	58	21*
---	------	--------	----	-----

>=====<

4	-.06	-.0093	58	17*
---	------	--------	----	-----

Calculating JMLE Fit Statistics

>=====<

Time for estimation: 0:0:1.303

Output to C:\Albert Documents\ICELDA\ZOU825WS.TXT

ice02701_Sesotho2_Subtest5_Winsteps.xlsx

PERSON	518 INPUT	518 MEASURED	INFIT	OUTFIT
TOTAL	COUNT	MEASURE REALSE	IMNSQ	ZSTD OMNSQ
MEAN	8.9	32.0	-1.11	.46
P.SD	4.2	.0	.82	.19
REAL RMSE	.49	TRUE SD	.65	SEPARATION 1.33

PERSON RELIABILITY .64

ITEM	32 INPUT	32 MEASURED	INFIT	OUTFIT
TOTAL	COUNT	MEASURE REALSE	IMNSQ	ZSTD OMNSQ
MEAN	143.3	518.0	.00	.11
P.SD	39.0	.0	.44	.01
REAL RMSE	.11	TRUE SD	.42	SEPARATION 3.91

ITEM RELIABILITY .94

Output written to C:\Albert Documents\ICELDA\ZOU825WS.TXT

CODES= 01

Measures constructed: use "Diagnosis" and "Output Tables" menus

The last subtest whose items were reduced was subtest 6 which focuses on grammar and text relations. The analysis for this subtest is displayed in Table 6.24. For this subtest, the same principles that were used for reducing items for subtests 1-5 were used and once more, *Winsteps* analyses (as shown in the convergence table) were used to augment reasons for reducing the items. Again, the specifications for selecting items for this subtest are displayed in Table 6.23.

Table 6.23: Specifications for subtest 6

Subtest	Specifications for 60 item test including number of items selected for each measured ability
Grammar and text relations	(12) The text is made up of grammatically related aspects – a range of components are assessed at the same time.

Table 6.24: NC-SES-TALA TiaPlus report for items making up subtest 6

Persons: All persons

Subtest(6): Grammar & text relations

146-187

Test and Item Analysis

		% #																			
Item	Item	----- P- and A- values -----										Mis- ----- Weighted -----									
Label	nr.	Weight	Key	A	B	C	D	E	F	O/D	sing	Max	Mean	P	Sd	RSK	[Lo	Rit	Hi]	Rir	AR
1	146	1	D	14	30	37	11*	0		8 42		1 0,11	11 0,32	0,32	3 10	17	2	54			
2	147	1	A	11*	30	29	23	0		8 41		1 0,11	11 0,31	0,31	0 7	13	-1	54			
3	148	1	B	17	24*	31	18	0		10 52		1 0,24	24 0,43	0,43	5 13	21	2	54			
4	149	1	D	22	27	25	17*	0		9 46		1 0,17	17 0,38	0,38	2 9	16	0	54			
5	150	1	A	10*	29	39	13	0		8 42		1 0,10	10 0,30	0,30	6 15	23	7	53			
6	151	1	B	13	31*	31	16	0		9 49		1 0,31	31 0,46	0,46	3 11	19	-1	54			
7	152	1	A	16*	24	30	20	0		9 48		1 0,16	16 0,37	0,37	6 14	21	4	54			
8	153	1	B	19	24*	30	18	0		9 47		1 0,24	24 0,43	0,43	10 18	25	7	53			
9	154	1	C	21	23	30*	16	0		10 50		1 0,30	30 0,46	0,46	20 28	35	16	52			

10	155	1	C	11	31	32*	16	0	9	45		1	0,32	32	0,47	0,47	16	24	31	12	53
11	156	1	A	11*	26	35	18	0	9	49		1	0,11	11	0,32	0,32	15	25	35	17	52
12	157	1	C	12	31	33*	14	0	9	49		1	0,33	33	0,47	0,47	27	33	39	22	52
13	158	1	C	18	22	33*	17	0	10	50		1	0,33	33	0,47	0,47	23	30	37	19	52
14	159	1	C	29	23	30*	12	0	7	35		1	0,30	30	0,46	0,46	21	28	35	17	52
15	160	1	C	11	31	37*	15	0	6	29		1	0,37	37	0,48	0,48	11	19	27	7	53
16	161	1	C	12	31	37*	12	0	8	39		1	0,37	37	0,48	0,48	17	24	32	12	53
17	162	1	A	16*	31	29	17	0	7	37		1	0,16	16	0,36	0,36	12	22	31	13	53
18	163	1	C	20	22	32*	17	0	9	45		1	0,32	32	0,47	0,47	28	34	40	23	51
19	164	1	C	21	23	32*	16	0	8	41		1	0,32	32	0,47	0,47	19	26	33	14	52
20	165	1	C	15	30	31*	17	0	8	40		1	0,31	31	0,46	0,46	25	32	38	20	52
21	166	1	B	12	32*	30	17	0	9	47		1	0,32	32	0,47	0,47	12	20	27	8	53
22	167	1	B	13	32*	30	17	0	8	43		1	0,32	32	0,47	0,47	13	21	28	9	53
23	168	1	B	18	25*	29	15	0	13	65		1	0,25	25	0,43	0,43	23	30	36	19	52
24	169	1	D	22	24	29	16*	0	9	48		1	0,16	16	0,36	0,36	18	26	35	18	52
25	170	1	D	10	29	35	16*	0	10	51		1	0,16	16	0,37	0,37	16	25	34	16	52
26	171	1	B	14	29*	33	15	0	10	50		1	0,29	29	0,45	0,45	18	26	33	15	52
27	172	1	A	15*	30	30	15	0	10	51		1	0,15	15	0,36	0,36	17	26	35	17	52
28	173	1	A	17*	24	31	16	0	12	63		1	0,17	17	0,38	0,38	-9	-2	5	-12	55
29	174	1	B	21	25*	28	17	0	10	50		1	0,25	25	0,43	0,43	20	28	35	17	52
30	175	1	D	10	31	33	16*	0	9	49		1	0,16	16	0,37	0,37	14	23	31	14	53
31	176	1	B	13	33*	32	13	0	9	45		1	0,33	33	0,47	0,47	23	30	37	19	52
32	177	1	C	14	26	32*	19	0	8	43		1	0,32	32	0,47	0,47	24	31	38	20	52
33	178	1	B	18	24*	30	17	0	11	57		1	0,24	24	0,43	0,43	23	30	37	20	52
34	179	1	C	24	24	27*	15	0	11	58		1	0,27	27	0,44	0,44	17	25	32	14	53
35	180	1	C	10	29	37*	15	0	10	52		1	0,37	37	0,48	0,48	18	25	32	13	53
36	181	1	D	13	32	33	13*	0	10	51		1	0,13	13	0,34	0,34	7	16	25	8	53
37	182	1	A	14*	31	30	16	0	9	47		1	0,14	14	0,34	0,34	13	23	32	14	53
38	183	1	B	21	21*	28	18	0	12	64		1	0,21	21	0,41	0,41	19	27	34	17	52
39	184	1	A	32*	20	13	13	0	22	113		1	0,32	32	0,47	0,47	15	23	30	11	53
40	185	1	C	11	26	29*	12	0	23	117		1	0,29	29	0,45	0,45	13	22	30	10	53
41	186	1	D	21	21	22	14*	0	22	115		1	0,14	14	0,35	0,35	8	16	24	7	53
42	187	1	A	21*	20	18	17	0	24	125		1	0,21	21	0,40	0,40	5	14	22	3	54

```

-----
SubGroup number      : 0          SubTest number      : 6
Number of persons in test : 519      Number of selected items : 42
Minimum test score    : 0          Maximum test score   : 42
Average test score    : 10,20      Standard deviation    : 3,94
Average P-value       : 24,30      Std. Error of Measurement : 2,69
Average Rit           : 0,22
Coefficient Alpha      : 0,53      SE Coeff. Alpha      : 0,03
GLB                   : 0,75      Asymptotic GLB coef  : Unavailable
Items used in GLB proc : 42

```

90% Confidence limits for Coefficient Alpha: (0,48 =< 0,53 =< 0,58)

Estimated Coefficient Alpha if this test had a standard norm length of 40 items: 0,52 (Spearman-Brown)
Cito, Measurement and Research Department. Arnhem, the Netherlands. © 2013.

The *Winsteps* convergence table (6.24) shows that the item reliability was acceptable at ,95 and the person reliability was lower than expected parameters at 6,3. The person reliability was brought down by 11 persons that were removed by the model (reducing the number from 518 to 507). However, after two iterations, the program was able to produce infit and outfit measures that were well within the required

parameters. They were both at 1,00 and 1,01 respectively and this indicated that items and persons used in this analysis were fit for this test.

Table 6.25: NC-SES-TALA subtest 6

```
518 PERSON Records Input
Building Index ...

CONVERGENCE TABLE
-Control: Sesotho2_Subtest6_Winsteps.txt Output: \ICELDA\ZOU224WS.TXT
| PROX      ACTIVE COUNT      EXTREME 5 RANGE      MAX LOGIT CHANGE |
| ITERATION PERSON  ITEM CATS    PERSON  ITEM    MEASURES STRUCTURE|
>=====<
|   1   518   42   2    3.87  1.46   -3.7136      |
>=====<
|   2   507   42   2    4.05  1.54   -1.1712      |
Index Size: 25 with 1 Patterns
Probing data connection: to skip out: Ctrl+F - to bypass: subset=no
Processing unanchored persons ...
>=====<
Data fully connected. No subsets found
|Control: Sesotho2_Subtest6_Winsteps.txt Output: \ICELDA\ZOU224WS.TXT
|  JMLE    MAX SCORE  MAX LOGIT   LEAST CONVERGED  CATEGORY STRUCTURE|
| ITERATION RESIDUAL*  CHANGE PERSON  ITEM  CAT  RESIDUAL  CHANGE|
>=====<
|   1     -2.13   -.3022   91   35*      |
>=====<
|   2     -.58   -.0900    1   35*      |
>=====<
|   3     -.15   -.0267   91   35*      |
>=====<
|   4     -.04   -.0079   91   16*      |
-----
Calculating JMLE Fit Statistics
>=====<
Time for estimation: 0:0:1.627
Output to C:\Albert Documents\ICELDA\ZOU224WS.TXT
ice02706_Sesotho2_Subtest6_Winsteps.xlsx
-----
| PERSON    518 INPUT    518 MEASURED      INFIT      OUTFIT  |
```

```

|      TOTAL    COUNT    MEASURE REALSE    IMNSQ    ZSTD OMNSQ    ZSTD|
| MEAN    10.2    42.0    -1.32    .42    1.00    .1    1.01    .1|
| P.SD     3.9     .0     .78    .22    .07    .5    .16    .7|
| REAL RMSE .47 TRUE SD  .61 SEPARATION 1.29 PERSON RELIABILITY .63|
|-----|
| ITEM    42 INPUT    42 MEASURED          INFIT          OUTFIT  |
|      TOTAL    COUNT    MEASURE REALSE    IMNSQ    ZSTD OMNSQ    ZSTD|
| MEAN   125.9   518.0     .00    .11    1.00    .0    1.01    .0|
| P.SD   43.6     .0     .53    .02    .03    .7    .07    1.1|
| REAL RMSE .11 TRUE SD  .51 SEPARATION 4.57 ITEM RELIABILITY .95|
|-----|

```

Output written to C:\Albert Documents\ICELDA\ZOU224WS.TXT

CODES= 01

Measures constructed: use "Diagnosis" and "Output Tables" menus

The above process followed by the panel of experts enabled them to select a total number of 60 items for a refined NC-SES-TALA. These 60 items were then included in an *Excel* data sheet similar to the original data sets (which contained all test items) for a further analysis of the now refined NC-SES-TALA (see appendix A3).

6.2.5 Analysis of the results of the refined NC-SES-TALA

The initial research aim for this study primarily focused on selecting either the T-SES-TALA or the NC-SES-TALA as the most adequate and appropriate instrument for usage as part of Grade 12 examinations. It should be clear from the preceding discussions that at this stage of the analysis, results point to the NC-SES-TALA as the most adequate and appropriate for this purpose. Following the above premise, an analysis of the results for the refined test will be provided below.

Under this subheading – “analysis of the results of the refined NC-SES-TALA” a graphical determination is displayed in Figures 6.3 – 6.7 (scatterplots), and numerically in Table 6.27 to illustrate dimensionality of the test through *TiaPlus* analysis (Cito, 2013:20). All subtests displayed in all scatterplots from 6.3 to 6.7 show satisfactory profiles as all of them show either one outlier or none at all. Moreover, for the intercorrelations between subtests all those highlighted in Table 6.26 are acceptable

between the values of 0.2 and 0.5. According to Van der Walt and Steyn (2007), these figures indicate that these subtests test different aspects of the construct.

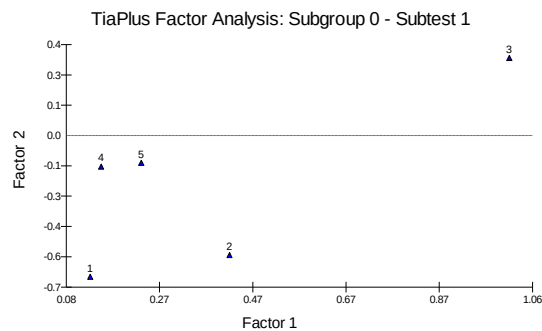


Figure 6.3: TiaPlus factor analysis of scrambled text for refined NC-SES-TALA

The same inference can be made with regard to subtests 2, 3, 4, and 5 as can be seen from the scatterplots displayed below in Figures 6.4, 6.5, 6.6, and 6.7. Although slightly sparsely spread, items occupy both the right upper quadrant and the right lower quadrant. This is different from the inclusive overall spread which is spread all over the 4 different quadrants of the NC-SES-TALA. The reason for the distribution in the overall scatterplot will be further discussed below.

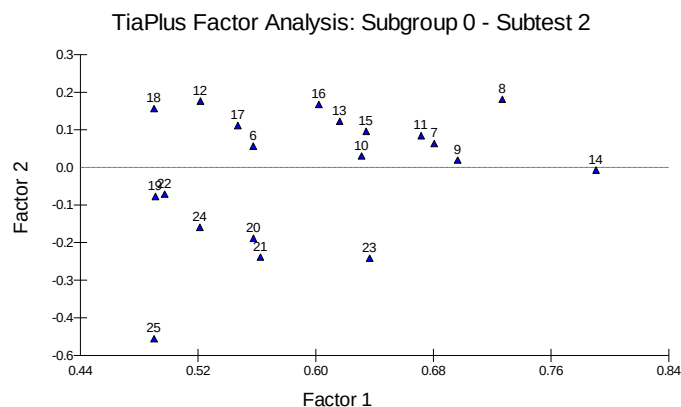


Figure 6.4: *TiaPlus* factor analysis of text comprehension for refined NC-SES-TALA (Ref-SES-TALA)

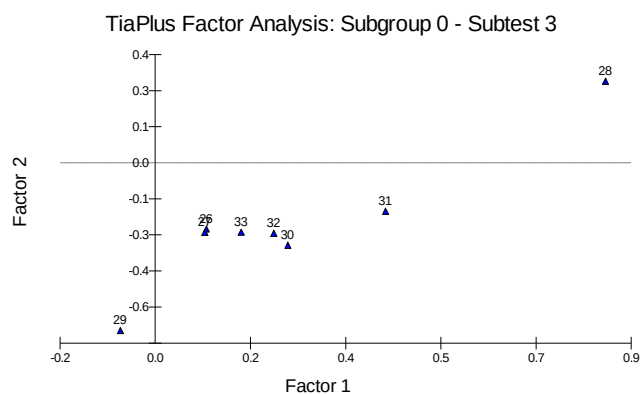


Figure 6.5: *TiaPlus* factor analysis of interpreting graphs and visual information for refined NC-SES-TALA

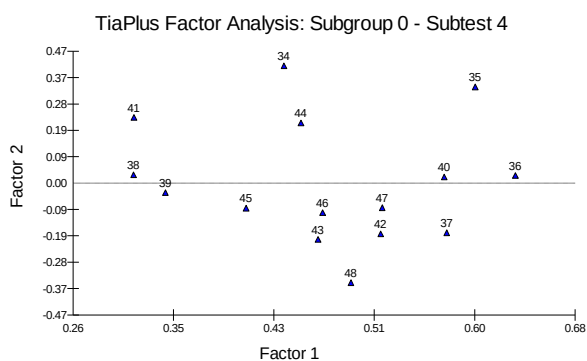


Figure 6.6: *TiaPlus* factor analysis of vocabulary for refined NC-SES-TALA

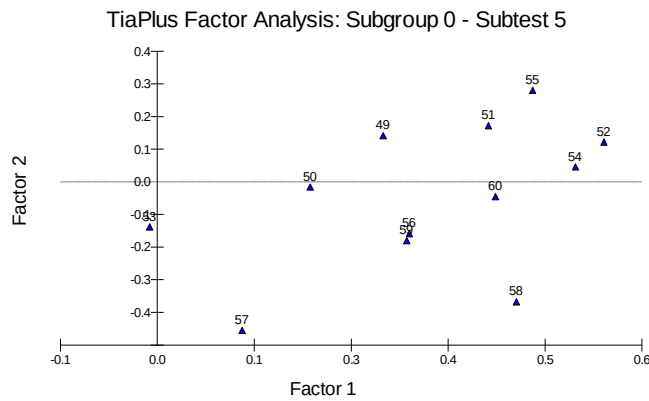


Figure 6.7: TiaPlus factor analysis of grammar and text relations for refined NC-SES-TALA

It is further necessary to mention that, based on the subtest intercorrelation reflected in Table 6.25, the difference between the performance of the original 187 item NC-SES-TALA (Table 6.9) and the refined NC-SES-TALA, is minimal. The subtest intercorrelation for the original 187 items NC-SES-TALA was better for two subtests when compared to the refined 60-item NC-SES-TALA. This was to be expected as the removal of one of the two scrambled texts from NC-SES-TALA that had a better combined performance by replacing them with the one that performed better between the two, would present a slightly weaker scrambled subtest because of it's value. It should be remembered that the second scrambled text that was chosen was selected on thematic condition only and not so much on the strength of the items. Nevertheless, the p-values and Rit values for this difference corroborate this evidence. How each of these (p-value and Rit value) contribute to the productivity of the test, was discussed in detail above in "detailed analysis of the data". These two concepts are used to explain what the facility value of the test and the discrimination index of the test are. What one observes from the difference between these concepts based on the analysis by *TiaPlus* is that the NC-SES-TALA has an average p-value of just 0,25, which indicates that the test was difficult. Moreover, the average p-value for the Refined Newly Created Sesotho TALA (Ref. NC-SES-TALA) is 0,29, which is slightly higher than for the NC-SES-TALA, but still rendering it difficult (cf. Table 6.26). Nevertheless, it should also be remembered that the average p-value for the ENG-TALA was 0,62.

This indicates that it was far more manageable for learners in comparison to NC-SES-TALA and its concomitant refined version. As already indicated, the desirable p-value for a test of this kind should be above 0,15, but below 0,84 (Steyn, 2018:57). However, with regard to p-values, one should also include an explanation of whether the test discriminated well or not in order to determine if the test is useful or not.

The discrimination index of a test of this kind should be above 0,25 (Steyn, 2018:57). For the NC-SES-TALA, the average rpbis produced by *TiaPlus* is 0,279 for the test as a whole and an average of 0,26 for the Ref. NC-SES-TALA (cf. Table 6.26). While this measure is above 0,25, which is commendable, it should also be noted that the average for the ENG-TALA as a whole is far beyond this at 0,317. It should also be noted again that the difference between the NC-SES-TALA and the Ref. NC-SES-TALA is not significant.

Table 6.26: NC-SES-TALA subtest intercorrelations for refined NC-SES-TALA

Persons: All persons		Items: All items				
Table of Subtest Intercorrelations						
	Total	Subtest(s)				
Subtest	test	1	2	3	4	5

--						
Scrambled text	1	0,19				
Text comprehens	2	0,78	0,12			
Interpreting gr	3	0,41	0,01	0,03		
Vocabulary	4	0,62	-0,05	0,21	0,24	
Grammar & text	5	0,35	-0,06	-0,06	0,26	0,11

Table 6.27: Summary statistics of NC-SES-TALA and Ref. NC-SES-TALA

NC-SES-TALA					Ref. NC-SES-TALA			
	Items	Mean	Mean P	Mean rpbis	Items	Mean	Mean P	Mean rpbis
All Items	187	52.20	0.279	0.17	60	18.57	0.310	0.26
Scored Items	187	52.20	0.279	0.17	60	18.57	0.310	0.26
Scrambled test 1	5	1.13	0.226	0.47	5	0.98	0.197	0.51
Scrambled test 2	5	0.98	0.197	0.51	-	-	-	-
Text Comprehension	73	23.45	0.321	0.31	20	7.38	0.369	0.51
Interpreting graphs and visual information	30	7.59	0.253	0.22	8	2.48	0.310	0.42
Vocabulary	32	8.85	0.277	0.30	15	4.21	0.281	0.43
Grammar and text relations	42	10.20	0.243	0.22	12	3.52	0.293	0.39

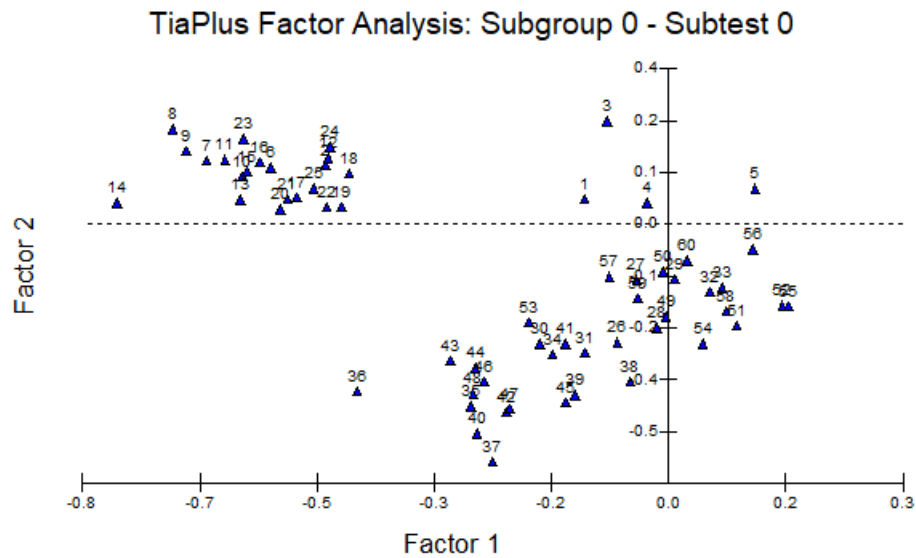


Figure 6.8: *TiaPlus* factor analysis of the complete, refined 60 items for the NC-SES-TALA

The *TiaPlus* factor analysis of the complete refined NC-SES-TALA as represented in Figure 6.8 presents a different picture when compared to the individual subtests. The individual scatterplots indicate items lying at the upper and the lower right hand side of the quadrant showing that they actually test multidimensional-related abilities. However, the scatterplot for the complete new refined 60-item NC-SES-TALA reflects items positioned in all quadrants of the scatterplot. Items 8 to 25 are situated in the first left quadrant of the scatterplot, 1 to 5 in the second right quadrant, 26 to 57 in the left lower quadrant, and 58 to 60 in the right lower quadrant. The interesting factor regarding this distribution lies in the proximity of each group of items. Items 1 to 5, 26 to 57 and 58 to 60 lie within close proximity of each other, while items 8 to 25 might be viewed as outliers because of their distance from the other items. Those lying within the same proximity with each other, could imply homogeneity of items and proof that the test construct is multidimensional or rich in nature (Weideman, 2017:36). This is desirable since, similar to the TALL, items clustered together prove that they measure a similar trait, but these clusters spread in groups indicate the multidimensionality of the test (Du Plessis *et al.*, 2012:126). Moreover, similar to what Van der Slik and Weideman (2005) state in the case of TALL, an argument could may be made in the

case of TALA that this spread is proof that the test measures different aspects of high knowledge and high skill of language and should not be reduced to a single homogenous concept (Du Plessis *et al.*, 2012:126). The construct of a test which displays such attributes is rich and measures different aspects of a single domain (which is not a discreet domain).

The reason for the difference in the scatterplots between the inclusive scatterplot (Figure 6.8) and individual subtest scatterplots (Figures 6.3 to 6.7) could be caused by the outcome of the entirety of the test as compared to the outcome of each subtest considered individually (i.e., the sum of the whole is always better than sum of the parts). Items of subtests are different from each other with regard to homogeneity – which is what factor analysis is supposed to measure if one compares only a few items, for example, five items of the same type in the scrambled text. However, when one compares potentially disparate measures, for example, one subtest with 59 other items, and a group of potentially different characters, there is expectably greater variation. It is a reasonable expectation not to expect the congruency between the items within the complete refined 60-item test to be significantly higher. This is caused by the reliability level of the NC-SES-TALA that would influence the reliability of the 60-item test. A test of such low averages – indicating that it is extremely difficult for its population – would not be expected to give the desired outcome in that regard.

6.3 Formulating a validity argument for the NC-SES-TALA

The validity argument that is presented in this section will address issues that relate to the adequacy and appropriateness of the NC-SES-TALA as the most relevant test (compared to the T-SES-TALA) for use in Grade 12 Sesotho HL examinations, based on the research questions the study set out to answer. An integrated approach will be used in the formulation of this validity argument, i.e. the contemporary perspective and argument-based perspective to validity will be integrated as stated in Chapter 4 (4.2.3). As mentioned in Chapter 4, the contemporary perspective will include both the constitutive as well as the regulative conditions for validation. For the argument-based approach, principles advocated by Chapelle (2011:19) would be incorporated.

Based on Kane's (2006) support of argument-based reasoning, scores as reflected above, show that NC-SES-TALA is a more adequate and appropriate assessment tool when compared to the T-SES-TALA. This is reflected by the different kinds of internal validity, i.e. construct validity, face validity, and content validity; as well as by external types of validity which include predictive validity, consequential validity, and concurrent validity. All these assertions will be elaborated upon in the following section.

Construct validity (as part of internal validity) in this argument is partly reflected by reliability scores and supported by the theoretical justifications outlined about item specifications for each subtest. The reliability scores reflected in the data analysis of the NC-SES-TALA by *TiaPlus* show that such scores are higher than the norm, i.e., the accepted alpha of 0,8 for the whole test (Van der Walt & Steyn, 2008:196). Construct validity is further reflected by the scores between different subtests which are between 0,2 and 0,5 which may as well indicate that the test measures different aspects of the construct (Van der Walt & Steyn, 2008), as explained above in 6.2.3. Correlations between subtests and the whole test also confirm that the NC-SES-TALA show higher reliability measures than the T-SES-TALA. This issue of scores between subtests and the whole test is also an indication that different aspects of the construct are tested as per the reliability measure (alpha requirements) for construct validity for the whole test and different attributes are tested for different subtests (Van der Walt & Steyn, 2007: 42).

Face validity is another type of internal validity which should be discussed in this argument as was mentioned in the introduction to this chapter. There are numerous definitions of face validity, but two definitions that can be objectively applied to TALA will be briefly discussed to illustrate how TALA conforms to face validity. According to Bachman (1990:285), face validity is the appearance of the relationship between the test and an external criterion, and according to Weideman (2020:60), face validity is based on ensuring effective measurement by using adequate test instruments or assessment material. When one examines specifications for tasks as contained by each subtest of TALA, one realises that TALA represents reality or an external criterion as explained by Bachman (1990:285). It is an assessment instrument that is used to

evaluate a real-life situation. It is used to assess language proficiency and achievement of Grade 12 learners who are doing Sesotho HL. It is a reflection on a real life domain which is proficiency in a target language which is Sesotho. This definition of face validity combines the definition of Bachman (1990:285) and that of Weideman (2020:60). This notion of TALA agrees with the assertion by Steyn (2013:93) that TALA has high face validity. Steyn (2013:93) bases her assertion on the overall judgement that was done by a panel of evaluators for her study. They found the TALA to be measuring the components of the construct that it was intended to measure (Steyn, 2018: 93). However, Steyn also cautions against being too dogmatic about this assertion as further research is necessary for it to be fully proclaimed.

For content validity as part of internal validity, the appropriateness of the TALA was elucidated in Chapter 4 (4.2.2) and Chapter 5 (5.3.2). It was highlighted that both the content coverage and relevance of TALA meet the required standard of this validity measure. Since the Sesotho TALA followed the same design, principles, parameters, and construct as the ENG-TALA, it is reasonable to assume that its content validity meets the same standard as the ENG-TALA.

Criterion-related validity as part of external validity was discussed in detail in Chapter 4 (4.2.1). It was stated there that criterion-related validity can be split into both concurrent and predictive validity. Concurrent validity for the two Sesotho TALAs implies a comparison between the T-SES-TALA and the NC-SES-TALA. The comparative analyses of scores between these TALAs indicate that the NC-SES-TALA seems to be more adequate and appropriate than the T-SES-TALA (4.2.1). This is evidenced by a NC-SES-TALA with a higher reliability of 0,82 compared to the 0,41 for T-SES-TALA. With regard to subtest intercorrelation, this is evidenced by six instances of intercorrelation scores of the NC-SES-TALA which are within the acceptable parameters of between 0,2 and 0,5, while the T-SES-TALA has only one instance of intercorrelation within this bracket. The T-SES-TALA intercorrelation within this bracket is 0,18 between vocabulary, and grammar and text relations. As mentioned in Section 6.2.3, Table 6.9, the NC-SES-TALA intercorrelations within this bracket are 0,18 between scrambled text 1 and text comprehension, 0,15 between

text comprehension and scrambled text 2, 0,16 between vocabulary and text comprehension, 0,37 between vocabulary and interpreting graphs and visual information, 0,19 between grammar and text relation, and interpreting graphs and visual information, and 0,30 between grammar and text relations, and vocabulary. Additional evidence to support the notion that the NC-SES-TALA is more adequate and appropriate than the T-SES-TALA is based on the correlation between the subtests and the test as whole. According to Van der Walt and Steyn (2008:196), the correlation between subtests and the whole test should be around 0,7. The notion that the NC-SES-TALA performs better than the T-SES-TALA is also supported by the fact that 3 subtests of the T-SES-TALA score above 0,5, 1 is below 0,5, 2 are equal to 0,2, while for the NC-SES-TALA, 1 is above 0,7, 1 above 0,5, 2 above 0,4, 1 is equal to 0,21, and 1 below 0,16.

Predictive validity, on the other hand, implies that the test would be able to predict future performance and behaviour of learners. As it stands with regard to the NC-SES-TALA, the average performance for this refined test is very low at 31%, which indicates that the test was arguably too difficult for this population. It, therefore, might be a difficult component to add to the Grade 12 Sesotho HL 20% as proposed by this study. A comparable test that was measured against the NC-SES-TALA was the T-SES-TALA. Clear areas of differences between the two tests were established and based on these one could argue that the NC-SES-TALA performs better than the T-SES-TALA. It is established what predictions can be expected if a test like this is applied for further investigations. It was also established that ENG-TALA is a better predictor of outcomes than the NC-SES-TALA because of the level of difficulty of the NC-SES-TALA.

As part of external validity, it is further important to determine to which extent the NC-SES-TALA meets the requirements for consequential validity. As stated in Chapter 4 (4.2.2.1) consequential validity deals with issues of bias and potential sources of bias, especially in terms of objective and subjective scoring. The analysis of results indicated, by means of the P-value, the extent to which different items reflected bias to different learners in this study. In order to overcome bias of gifted learners being

asked questions which are too easy for them, and weaker learners being asked questions that are too difficult for them, the Rit and the P-values of items selected for the refined NC-SES-TALA were carefully considered to accommodate both cohorts of learners. However, with regard to bias in terms of cultural background, background knowledge, and ethnicity, it can in summary be stated that learners used in this study come from a poor socio-economic background, specific ethnic group, and cultural background, so data might be skewed due to its homogeneity. Their place of residence is characterised by a housing backlog, inadequate physical and social infrastructure, low resource allocation, a high level of illiteracy and a poor skills base among a large part of the population, and income as well as infrastructure disparities. The four schools that were evaluated in this research belong to a third level of poverty (Quintile 3) according to the DBE's classification of schools (where 1 is the poorest and 5 the richest). Quintile 3, according to the description given by the DBA, appears average, which is supposed to imply that their socio-economic status is just average and free from setbacks brought about by it, but when one considers that schools belonging to Quintile 3 do not pay school fees, include a free feeding scheme and draw pupils from mostly informal settlements, one can only conclude that the classification does not reflect the true socio-economic situation of these learners. In fact, Quintile 3 is classified with Quintiles 1 and 2 (Urgorji & Francis, 2019:114) which indicates that it belongs to the poor level regarding socio-economic status. The home language academic performance of these learners is reflected in Table 6.24. The table indicates their overall pass rate in the NSC from 2016 to 2020 in the Sesotho HL subject. The report that provides further explanation of this table states that there was a decline in Sesotho HL performance for the past five years (2016–2020) (DBE, 2020:6). What is interesting is that for the past five years (from 2016–2020), the average academic performance of learners writing the Grade 12 Sesotho HL examination has been 99,32%, despite the decline. The average academic performance of learners who wrote Grade 12 English HL, on the contrary, remained stable at 93,28% during that five-year period as reflected in Table 6.25. The cohort of learners that were used for the ENG-TALA (Steyn, 2018:54) is similar to that reported on in the DBE's diagnostic report (DBE, 2020:6) (cf. Table 6.25). This cohort of learners belongs to former Model

184

C schools – most of these schools were previously advantaged schools. It should, however, be noted that according to Steyn’s study, this cohort also included learners from disadvantaged backgrounds. However, most learners from former Model C schools come from advantaged backgrounds. Historically, learners from former Model C schools always performed better than learners from former DET schools overall. It is interesting, therefore, that when one compares the Tables (6.28 and 6.29) based on HL performance, the performance in Sesotho HL is better than that of English HL. One of the reasons provided by Du Plessis (2017: 35) for this state of affairs, is the lack of variation in questions belonging to Indigenous African HL examination papers as questions are repeated every year under the guise that there are not enough literature texts available written in the Indigenous African languages. The high academic achievement for the Indigenous African languages (and not for other subjects) is an indication that Sesotho HL Grade 12 results are not a good predictor of whether or not learners are likely to pass the NSC overall (DBE, 2014:178), as explained in Chapter 2. A study by Ogbonnaya and Awuah (2019) further indicates that learners from Quintile 4 schools achieved significantly higher scores than learners from lower quintile schools in the achievement scores at all levels of Bloom’s taxonomy, except synthesis. It is thus not surprising that tests such as the TALA would yield better results from former Model C learners than learners from former DET schools, as has been clearly shown by the results of the SES-TALA compared to those of the ENG-TALA. It is therefore advisable to enlist the use of a test that can be statistically analysed to equalise Indigenous African HL with English HL according to the strategy suggested by this study.

Table 6.28: Academic performance of learners in Sesotho HL from 2016–2020 (DBE, 2020:153)

YEAR	N who wrote	N who achieved at 40% and above	% achieved at 40% and above
2016	32 198	32 002	99,4
2017	30 776	30 583	99,4
2018	31 168	30 985	99,4
2019	32 509	32 270	99,3
2020	32 104	31 827	99,1

Table 6.29: Academic performance of learners in English HL from 2016–2020 (DBE, 2020:153)

YEAR	N who wrote	N achieved at 40% and above	% achieved at 40% and above
2016	107 967	101 610	94,1
2017	105 705	98 362	93,1
2018	106 577	98 823	92,7
2019	108 483	99 843	92,0
2020	110 784	104 724	94,5

In the final analysis, when one considers the validity argument with regard to both the internal and external validity of the T-SES-TALA and the NC-SES-TALA presented above, there is a strong argument that the NC-SES-TALA can be used with higher

confidence than the T-SES-TALA. This better alignment of performance in favour of the NC-SES-TALA, is evidenced by the overall reliability differences between the two tests, the subtest intercorrelation differences between subtests that make up each of the two tests, and the reliability differences between subtests and the whole test.

Another area of focus with regard to an evaluation of the validity of the NC-SES-TALA, is the difference between the unrefined 187-item NC-SES-TALA and the 60-item refined NC-SES-TALA. In this regard, there were small pockets of improvement displayed by the 60-item refined NC-SES-TALA as compared to the unrefined 187-item NC-SES-TALA. These pockets of improvement indicate that, in addition to other principles of responsible design, refining a test is a wise strategy to follow for designing a language test.

The pockets of improvement regarding internal validity were displayed in the following areas:

- For construct validity, one major advantage was displayed by the scoring validity or reliability in terms of 0,8 for the whole test and the scores of between 0,2 and 0,5 between the subtests (cf. Table 6.9) for face validity. Specifications and the number of items measuring those specifications for each subtest illustrate and confirm the appropriateness of this instrument to be used as one of the TALA tests.
- For content validity, Chapter 4 and 5 were able to show appropriate content coverage and relevance of the TALA.
- For criterion validity, concurrent score analysis between the T-SES-TALA and NC-SES-TALA was able to determine that the NC-SES-TALA was the better performing of the two tests.

In addition to the above improvement, there were also small pockets of improvement with regard to areas of external validity:

- For consequential validity, in order to overcome bias, the Rit and the P-values of items selected for the refined NC-SES-TALA were carefully

selected to accommodate both cohorts of learners as displayed in the subtest tables. It was also explained why the ENG-TALA might have performed far better than SES-TALA in terms of test-takers' socio-economic background.

These pockets of improvements for a refined NC-SES-TALA, in combination with the *a priori* considerations for validation such as item specifications and construct defence all favour a recommendation of the refined NC-SES-TALA.

In order to conclude this argument successfully, one needs to address the issue of the hypothesis that was stated in the introduction to this study, namely that if one TALA between the Translated Sesotho TALA and the Newly Created Sesotho TALA has better internal and external validity than the other, the TALA with a better internal and external validity shall be considered a better TALA. Consequently, a better TALA shall be recommended to form 20% of the 300 marks for Grade 12 home language examination paper because of its better validity. As experienced in the deliberations leading to this stage, it was observed that the Newly Created Sesotho TALA had a better internal and external validity as compared to that of the T-SES-TALA. For this reason, the NC-SES-TALA is recommended to form 20% of the 300 marks for Grade 12 HL examination paper.

6.4 Conclusion

This chapter demonstrated how the empirical research results were arrived at using the procedure that was proposed in Chapter 1 of this study. As proposed in the beginning, the central theoretical statement of this study, the hypothesis together with the research questions were used for outlining the process that was followed for the data analysis in the study. This process entailed the arrangement of raw data into data sets applicable for analysis by *TiaPlus* and *Winsteps*. This analysis revealed the status of each of the two SES TALA tests in terms of which of the two Sesotho tests was

more adequate and appropriate to be potentially used as 20% of the Grade 12 HL examination paper. This was corroborated by means of a validity argument that showed that despite being a difficult test, the NC-SES-TALA is the more appropriate test to be recommended for application in the Grade 12 HL examination paper. It was also outlined, based on the results presented above, that the data just met the required minimum in terms of parameters, thereby limiting the validity of this NC-SES-TALA as a conclusively valid test. The most probable explanation for this is the presence of confounding variables for which this study did not plan. Such variables include the fact that Sesotho is not used as a Language of Learning and Teaching (LoLT). Furthermore, even if Sesotho could be used as LoLT, other limitations such as a lack of vocabulary, resources and the ability of teachers to teach in this language could affect the results.

CHAPTER 7: REFLECTIONS, CONCLUSIONS AND RECOMMENDATIONS

7.1 Introduction

It was reported in Chapter 1 that the project to address the apparent inequality between HL examination papers was conceptualised in 2013 by ICELDA that was commissioned by the Council for Quality Assurance in General and Further Training (Umalusi) to investigate the matter of inequality.

Consequently, this study continued investigations in an attempt to achieve educational parity across Grade 12 HL examination papers. It was discovered that certain Grade 12 HL examination papers were on a higher difficulty level than others in terms of results that were produced at the end of each year. HL examination papers that were discovered to be more difficult were the Afrikaans and English papers.

Several reasons were advanced for the lack of educational equity mentioned in the paragraph above. These reasons were found to be based on the historical situation of apartheid, as well as the inadequacy of the process of reform and transformation that took place after the fall of apartheid in education.

In order to address the above-mentioned situation, ICELDA and Umalusi supported this study as a strategy to mitigate the situation. This strategy involved using an approach that would help to equalise different Grade 12 official HL examination papers. It was conceived that this would be achieved by affording all Grade 12 official HLs a fraction of a paper (20% of the 300 marks) that would conform to the same set of principles underlying language testing, thereby entrenching proper standards for these papers. This resulted in a framework where a fraction of the Grade HL examination papers (i.e., 20%) be based on a common construct of language ability, in order to achieve acceptable levels of reliability and to ensure validity as espoused by the Curriculum and Assessment Policy Statement of the Department of Basic Education.

In order to achieve the objectives set out above, this study adopted a methodology where one original test of academic literacy called the Test of Advanced Language Ability (TALA) in English was translated into one of the indigenous African HLs, i.e. Sesotho. Additionally, another test in Sesotho was designed anew in line with the same requirements of the original TALA. The aim of this, was to compare the performance of these tests, i.e. the Translated Sesotho TALA (T-SES-TALA) and the Newly Created Sesotho TALA (NC-SES-TALA) in terms of their reliability and validity, and other requirements of responsible test design. The one that would perform better, in terms of their statistical properties, would be recommended for inclusion in the 20% of the 300 marks of the Grade 12 Sesotho HL examination paper. This could further serve as a basis for other Grade 12 indigenous African official HL examination papers.

Based on the context outlined above, this chapter briefly summarises each of the main components of this study. This is done through a process of reflection on such components, followed by a conclusion outlining the implications of the empirical results of this study. The conclusion also incorporates the noted substantial issues emanating from the translation of texts versus developing tests anew. Limitations and values of the study are also considered. Lastly, this chapter outlines recommendations considered necessary to entrench the development of the TALA or a similar test of academic literacy for all the eleven official languages of the multilingual department of South African education system.

7.2 Summary of the main components of the study

It was learnt from this study that some of the key components of the study are based on the present understanding about both validity and reliability as main criteria upon which language testing is based. It was also learnt that the validity argument is important in explaining whether the experiment was a success or not. The validity argument that was framed was used to argue which between the T-SES-TALA and the NC-SES-TALA had higher equivalence with the ENG-TALA. The answer to this question was also based on the research question that was asked at the beginning of the study in Chapter 1.

It was discovered from literature that was reviewed that the present understanding about reliability and validity was mostly based on the notions of 'primacy of fairness', accountability, practicability, and responsible design. It was also discovered that reliability is not just viewed as a separate aspect of validity, but as part of the scoring validity (Shaw & Weir, 2007:7). The validity argument used an integrated approach between contemporary and argument-based perspective of validity to outline the argument. It did that by outlining reasons why the NC-SES-TALA had a higher equivalence with the ENG-TALA than the T-SES-TALA and how the conclusion to that effect was reached. Reasons advanced in that argument also addressed questions that were explicitly asked in the research question. Moreover, the answers to the research questions were also able to confirm that the NC-SES-TALA had a higher equivalence to the ENG-TALA than the T-SES-TALA. The validity argument outlined reasons why the NC-SES-TALA had higher equivalence and how the conclusion to that effect was arrived at.

7.3 Reflection on the translation of the Test of Advanced Language Ability (TALA)

In order to ensure that a test that would be produced from the translation of the English TALA complies with the rules, regulations and requirements of test translation, a number of established strategies and principles were followed as reported in the preceding chapters (particularly in Chapter 5, Section 5.3 - preparation of the stimulus material).

It was also reported that care was taken, first, to make sure that the construct of the original test was retained in the process of translation. This was accomplished by complying with the guidelines of the various requirements for a successful translation as espoused by translation bodies such as the International Test Commission (ITC) (2016). Second, all sections of the original template of English TALA together with its major elements were retained in the translated Sesotho version of the English TALA. As part of the strategy to ensure that the message was carried forward as a faithful version of the source text, back-translation was also employed.

In order to ensure that the above-mentioned strategies are in line with the requirements of the translation brief and the ITC, a team of expert translators and educators was used for translation and discussion of technical aspects of the text used (sources about translation procedures and strategies). This team also took part in a workshop on language testing before they started with the translation process.

7.4 Reflection on the creation of a new Sesotho TALA (NC-SES-TALA)

A new Sesotho TALA was created based on the same principles on which the original English TALA was created, i.e. it was created on the construct that was agreed upon by the project team. This construct was defined as “the assessment of 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 *et al.*, 2016). This construct was based on the notion of “high knowledge and high skill” as outlined for the TALA by CAPS. It is further encapsulated by abilities and tasks measured by this test as outlined in Chapter 5; Section 5.3.2 as well as Table 5.3 (with its concomitant explanations).

7.5 Reflection on the method of conducting an experiment for equalising English and Sesotho TALAs

In order to equalise the English and Sesotho TALAs, a quantitative methodology was utilised. In order to achieve the objective of equalising the English TALA and the Sesotho TALA, two Sesotho TALAs were developed, one a translated test and the other a newly created test. These two tests were then compared statistically to the English TALA as well as with one another. The Sesotho tests were marked electronically and statistically analysed for comparative purposes.

It was also reported in Chapter 5 that, according to the ethical code of conducting research with people, ethical clearance from certain authoritative bodies was obtained. Ethical clearance was granted by the NWU’s ECLM. In addition to this, permission to conduct research at certain identified schools and districts of the DBE was acquired

from the provincial office of the DBE in the Free State Province. Individual parents consent and learner assent were acquired from this population as well.

It was further reported in Chapter 5 that the experiment was conducted at four different secondary schools in the same township of Zamdela in Sasolburg, in the Fezile Dabi district. These schools offer English at FAL level and Sesotho at HL level. A total of 614 learners were expected to write the two Sesotho TALAs, but as reported, the numbers fluctuated according to learners who attended the test sessions each day. The total number of learners who ended up being tested ranged from a minimum of 514 to a maximum of 690 (cf. Tables 6.3 to 6.5).

The results of the T-SES-TALA and NC-SES-TALA were analysed using the *TiaPlus* and *Winsteps* statistical programs. The *TiaPlus* statistical program was utilised because both CTT analysis and Rasch analysis are employed to produce parameters for both internal and external validities. The benefit in making use of the *TiaPlus* software is that it provided both descriptive and inferential statistics about the test data. This was advantageous in that it provided information about the tests both at test and item level. *Winsteps* was chosen to augment the *TiaPlus* analyses so as to verify the parameters that were produced by *TiaPlus*. *Winsteps* provided two measures of reliability, namely, item reliability and person reliability. One could, therefore, discern at a glance whether the items gave an accurate picture of test takers' abilities and consequently, whether the scores of test takers provided a true picture of the level of difficulty of items (Green, 2013:167).

Based on the results of the analysis that was produced by both the *TiaPlus* and *Winsteps* statistical programs, it was explained in Chapter 6 that the better performing test was then selected for refinement. This refinement involved identifying those 60 items from NC-SES-TALA which showed the best productivity. This measure of productivity, based on scores as well as other validity evidence, was used to select items which could be included in the refined test. The refined test was again subjected to statistical analysis to produce results that would be used in the validity argument.

Consequently, a validity argument was formulated as a determination of which of the two tests, i.e. the T-SES-TALA or the NC-SES-TALA could be recommended for inclusion in the Grade 12 HL examination paper.

7.6 Reflection on the results of the Translated Sesotho TALA (T-SES-TALA)

The results of this experiment can be further refined and improved based on further research where gaps exist. Although the *a priori* validation of these tests relied heavily on the original template of the English TALA, the tests differed considerably with regard to their scoring validity. The overall reliability of the T-SES-TALA was unacceptably low (0,41) when compared to that of the ENG -TALA and the NC-SES-TALA which stood at 0,96 and 0,82 respectively. In addition, the intercorrelations between subtests of the T-SES-TALA were found to be negative and weak, which meant that the reliability between these subtests was low. It was further pointed out that the test score average was lower (50,53) in comparison to the ENG-TALA and the NC-SES-TALA that produced average test scores of 116,79 and 52,20 out of 187 respectively. Though the difference between the difficulty level of the T-SES-TALA and the NC-SES-TALA was marginal, the NC-SES-TALA remained the test with the higher equivalence to the ENG-TALA. Again, based on the frequency histogram (Figure 6.1) which was skewed to the left, there was indication that fewer learners got answers correct. The factor analysis (Figure 6.2) also showed many outliers, indicating that many items measured a second factor. All the factors mentioned above rendered the T-SES-TALA inadequate for the purpose of this study that sought to determine which of the two Sesotho TALAs was more appropriate and adequate for use.

7.7 Reflection on the results of Newly Created Sesotho TALA (NC-SES-TALA)

After the T-SES-TALA was eliminated for the reasons mentioned above, the NC-SES-TALA was subjected to a process of refinement. The purpose of this refinement was to improve the performance of the test by discarding items which did not perform as desirably as they should and using only items with better productivity for a final refined test (Myburg-Smit & Weideman, 2017:18). Several measures of analysis (as explained

in Chapter 6) were used to refine the 187-item NC-SES-TALA into a 60-item final version of the test.

As reported in Chapter 6, two statistical programs were again used for the process of refinement. Working on the assumption that validity measures other than scoring validity were in place, i.e. no additional validity processes were put in place before the 60-item NC-SES-TALA was analysed, the test was subjected to another *TiaPlus* statistical analysis. Items that were subjected to a second round of statistical analysis were identified making use of their P and Rit values and their combined performance and specifications. The overall reliability (Tables 6.3 to 6.5) and subtest intercorrelation (Table 6.3) reflected the difference between the performance of the 187-item NC-SES-TALA and the 60-item NC-SES-TALA. The difference between the reliability of the overall 187-item NC-SES-TALA (0,82) and the 60-item NC-SES-TALA (0.8) was only marginal. Moreover, based on Table 6.9 for the 187-item NC-SES-TALA subtest intercorrelation and Table 6.23 for the 60-item NC-SES-TALA subtest intercorrelation, both tests also had marginal differences in terms of the number of intercorrelations that did not meet the required parameters. However, based on the scatterplot for each subtest, it can be concluded that the refined test's subtests measured the factors required for each subtest (because there were no outliers in any of the scatterplots for these subtests). However, because of the poor performance of the learners to answer correctly, the average performance stood at only 31%. It should be noted that the average performance of the 60-item NC-SES-TALA was lower than that of the English TALA (0,96).

7.8 Reflection on the validity argument

It was reported in Chapter 6 that the validity argument presented in this study is based on both the internal and external validity of the test that would be selected as the most adequate and appropriate test between the T-SES-TALA and the NC-SES-TALA for the purpose of this study. Reflecting on the final results of the NC-SES-TALA mentioned above, one is able to surmise that, between the two tests, the NC-SES-TALA was the more adequate and appropriate test. The reflection was based on the

argument that was presented in Section 6.3 in Chapter 6. It was pointed out in that chapter that, based on its theoretical validity as well as scoring validity, the NC-SES-TALA could be considered the better of the two Sesotho tests.

The NC-SES-TALA could therefore be recommended for the hypothetical 20% of the 300 marks to be included in the Grade 12 HL examination paper.

Despite the fact that the T-SES-TALA was based on the ENG-TALA template and its accompanying validity and reliability features, the reasons for discarding this test in favour of the NC-SES-TALA after statistical analysis include:

- The overall reliability value of the NC-SES-TALA that was better than that of T-SES-TALA.
- The subtest intercorrelation reliability values of the T-SES-TALA that were found to be negative and weak as compared to those of NC-SES-TALA.

The design and the development of NC-SES-TALA was also based on the same template as well as the validity and reliability features of the ENG-TALA. However, its scoring validity ran short of reaching those of the ENG-TALA. As stated, the overall reliability value of the 187-item test stood at 0,8 and its subtest intercorrelations were reported in Table 6.9. The reliability value of the refined 60-item NC-SES-TALA was 0,77 and its subtest intercorrelations reported in Table 6.23.

7.9 Conclusions from the study

The main conclusions of this study are based on the analysis of the data and the specific issues included in the validity argument.

7.9.1 The level of difficulty of the test and its implications

The results of the NC-SES-TALA indicate that the test was difficult or the population for which it that was used was weak. As mentioned in Section 7.7, learners scored an average of 31% for the NC-SES-TALA. Since, according to the guidelines towards the design and development of a test of this nature, it is required that students be

familiarised with the test format prior to taking it, one of the reasons for this poor performance might be a lack of familiarity with this type of assessment by the test-takers (Du Plessis *et al.*, 2012:128). These learners were not exposed to a test in this format prior to the administration of this test. The reason for this was to avoid a negative washback effect where unfair inferences might be made by test-takers (Ward-Cox, 2020:129). However, after having administered the test and having analysed the test data, the need to familiarise learners with the test may outweigh the washback effect mentioned here.

Furthermore, the poor performance by the learners who wrote the test is surprising to some extent, considering that Sesotho HL is ranked as one of the subjects with a high pass rate at national level (see Tables 6.24 and 6.5). Another possible reason for the poor performance could have been learner apathy from certain members of this population as they might have considered this test, because of its format and content, as irrelevant and non-beneficial for their education. This notion is based on the premise that a certain number of answer sheets were found spoiled and not complying with the requirements of the questions, e.g. some answers that followed a patterned structure as explained in Chapter 6. Such answers were eventually removed from the data set.

7.9.2 The external and internal validity implications of the study

The external and internal validity of the NC-SES-TALA were discussed comprehensively in Chapter 6. It was posited that the NC-SES-TALA statistically performed better than the T-SES-TALA in terms of both external and internal validity. A refinement was subsequently done for the NC-SES-TALA. However, the difference in performance between the NC-SES-TALA and the Ref-NC-SES-TALA was minimal. However, the REF-SES-TALA still performed better than the T-SES-TALA in terms of subtest intercorrelation. This finding also provides an answer to the main research question of the study which sought to determine which of the two Sesotho TALAs performed best statistically. The better performance of the NC-SES-TALA regarding

internal and external validity measures clearly rendered this test the better performing of the two, albeit not necessarily a good one.

7.9.3 Translation versus creating tests anew: what do we learn?

As mentioned previously, both Sesotho tests that were used for this study (i.e. the T-SES-TALA and NC-SES-TALA) were found to be difficult for the population for which it was used. The average test score for the T-SES-TALA was 50,53 and 52,20 for the NC-SES-TALA. The statistical difference in average between the two tests was found to be minimal, or not significant. We learn from these averages that the difficulty level of these two tests was similar.

The results of this study further showed that apart from the validity measures discussed in 7.9.2, the newly created test performed better than the translated test on other aspects such as item facility, discrimination, and overall test reliability as well. The average facility value (i.e., the distribution of correct and incorrect answers) for the T-SES-TALA was 0,27 and for the NC-SES-TALA, 0,28, while the average Rit (which stands for item-total correlation, which distinguishes between a weak and a strong test taker) for the NC-SES-TALA was acceptable at 0,17 for item discrimination as opposed to the unacceptable value of 0,09 for the T-SES-TALA. The overall reliability for the whole test was acceptable at 0,82 for the NC-SES-TALA and unacceptable at 0,41 for the T-SES-TALA. This finding was unexpected since one would have expected better reliability of the translated Sesotho TALA, based on the high reliability measures reported for the English TALA from which it was translated. In addition, tests that utilised a similar process of translation as the one followed in this study reported better results in terms of internal and external validity than what is seen for the translated test.

7.9.4 Limitations of the study

One limitation of the study is the difficulty level observed for both the T-SES-TALA as well as the NC-SES-TALA. It could be discouraging for stakeholders such as learners, educators, and officials to make use of a test that is perceived to be too difficult for

learners. Perhaps more exposure for the test-takers to this type of assessment could improve their performance and subsequently the internal and external validity of the test. This exposure could be enhanced through the incorporation of tasks and its concomitant specifications in learning and teaching throughout the school calendar. This would help them to improve their comprehension in various texts used in different content subjects done at school.

Another limitation identified for this study relates to the specific population that was used for writing the tests. While the population that was used for the ENG-TALA was a combination of former Model C school learners and disadvantaged learners, this study used a population that was derived from former DET schools only. As argued in Chapter 6, these learners are regarded as poor (regarding socio-economic status) and disempowered to perform at a similar level to those learners that come from former Model C schools. According to Van der Slik and Weideman (2010), one of the contributing factors for the better performance by the English First Language group for the TALL was, in part, its heterogeneous socio-economic background especially by test-takers from a privileged socio-economic background (Van der Slik & Weideman, 2010:116). This utterance can be extrapolated to the population that took ENG-TALA if we consider that the TALL and the TALA share similar construct and structural features.

Another limiting factor with regard to this study is the issue of the Language of Learning and Teaching (LoLT) that is used by the population of this study. The LoLT for this population is English which is an additional language for all these learners. The LoLT for the population that was used for the ENG-TALA was also English, a HL for most of those learners. The population for this study, unlike the population used for the ENG-TALA did not have the advantage of receiving tuition in their HL (Sesotho) that was used for the TALA in their other school subjects. The former Model C learners therefore enjoy an added advantage of learning through their HL (English) with regard to content subjects as well as English HL as an independent subject. The former DET learners, however, learn Sesotho HL only as an independent subject. This could be,

in part, the reason why the population for this study did not perform as well as the sample for the ENG-TALA.

According to the contemporary perspective of validity as was discussed in Chapter 4, several authors advocated for the primacy of fairness as one of the most important principles of responsible test design. It is in this context that the issues discussed above should be considered as important limitations of the study.

7.9.5 The value of the TALA in the Sesotho HL examination paper

The value and relevance of this study lies in the guidance it provides with regard to creating equity between different HL examination papers in a multilingual education system. First, it explained comprehensively the methodology to be utilised in the design of a test of this nature. It further provided an explication of appropriate strategies and procedures which are necessary when using a similar approach for test design in an African language such as Sesotho. The way in which this was accomplished is subsequently summarised.

It was able to illustrate theoretically and practically how the translation of a test of this nature should be approached. Issues of approaches, strategies and procedures which are necessary to ensure the validity of a test of this nature were dealt with comprehensively. This includes matters such as the use of the guidelines of the ITC in order to ensure validity of the study such as those discussed in Chapter 5. The study also elaborated on the approach that was used in selecting an appropriate translation model (the Functional Approach) for a study of this nature (cf. Chapter 3).

In addition, this study also illustrated both theoretically and practically how issues of language testing should be handled for a test of this nature. This was done through the introduction of different perspectives of validity that were used in different time periods and culminated in the introduction of a contemporary perspective on validity (that was utilised in this study). This was concluded through an illustration of an appropriate way to construct a validity argument.

The study further pointed out limitations that could be explored by further research. This is important to ensure that difficulties encountered in this study are not repeated for future studies of this nature.

7.10 Recommendations

The notion of designing TALAs in all official South African languages could go a long way in establishing equity between Grade 12 HL examination papers. However, in order to promote this concept to the DBE, adequate academic publication is essential that would provide the necessary evidence of the potential value of the idea. More popularised reading material could further be distributed to schools and the general public in an attempt to elicit interest and support for the idea. Such material could be further used as potential training resources for teachers and officials. Such reading material should be available in all official languages of South Africa as a response to the true nature of multilingual teaching and learning in the country.

As mentioned in 7.9.4, the researcher recommends that test-takers be exposed to the format of the TALA (if possible, during ordinary class time) before they write such a test to mitigate for the potential influence that unfamiliarity with the format of the test may have on test scores.

It is envisaged that additional studies would be conducted in which TALAs would be designed for all the remaining official South African languages, and that through this process, it would be possible to equalise all Grade 12 HL examination papers. Studies such as this one, as well as those by Du Plessis (2017) and Steyn (2018), could be used productively by future studies to mediate issues of procedure for the development of a test of this nature.

REFERENCES

Alderson, J.C., Clapham, C., & Wall, D. 2005. Language test construction and evaluation. Cambridge: Cambridge University Press.

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 XVIlth World Congress of the International Federation of Translators held at Tampere, Finland, 4-7 August 2005.

Australian Medical Assessment Collaboration. (Not dated). Determining the quality of assessment items in collaborations: aspects to discuss to reach agreement. <https://www.acer.org/files/quality-determination-of-assessment-items-amac-resource.pdf>: Date accessed: 22/11/2021.

Babbie, E. & Mouton, J. 2004. The practice of social research. Cape Town: Oxford University Press.

Bachman, L.F. 1990. Fundamental considerations in language testing. Oxford: Oxford Press.

Bachman, L. & Palmer, A. 2010. Language assessment in practice. Oxford: Oxford University Press.

Bachman, L.F. & Palmer, A.S. 1996. Language testing in practice. Oxford: Oxford University Press.

Bachman, L.F. & Palmer, J.E. 2010. Language assessment: Gate-keepers or door openers? *The handbook of educational linguistics*. Oxford:Wiley – Blackwell. 456-468.

Bachman, L.F. 1990. Fundametal considerations in language testing. New York: Oxford University Press.

Bassnet, S. 2014. Translation: the new critical idiom. New York: Routledge.

Bassnett, S, & Trivedi, H. 2012. Postcolonial translation: theory and practice. London: Routledge.

Brown, J.D. 2012. Classical test theory. In Fulcher, G. & Davidson, F. (Eds), *The Routledge handbook of language testing*. New York: Routledge. pp. 323-335.

Butler, G. 2017. Translating the Test of Academic Literacy Levels into Sesotho. *Journal of Language Teaching*, 51(1). page numbers

Chapelle, C.A. 1994. Is a C-test valid for L2 vocabulary research? *Second language research*. 10. 157 – 187.

Chapelle, C.A. 1999: Validity in language assessment. *Annual review of applied linguistics*. 19: 254-272.

Chapelle, C.A. 2010. Does an argument-based approach to validity make a difference? *Educational measurement: issues and practice*: 29(1): 3-13.

Chapelle, C.A. 2011. Validity argument for language assessment: the framework is simple. *Language testing*: 29(1): 19-27.

Chapelle, C.A., Enright, M.E., & Jamieson, J. 2010. Does an argument-based approach to validity make a difference? *Educational measurement: issues and practice*, 29(1) 3-13.

Cito. 2013. TiaPlus user's manual. https://TiaPlus.cito.nl/TiaPlus_Users_Manual.pdf. Date of access: 22 Nov. 2021.

Cumming, A. & Berwick, R. (Eds). 1996. Validation in language testing. Clevedon, UK: Multilingual Matters *Ltd*. Cumming, A. & Berwick, R. (Eds). 1996. Validation in language testing. Clevedon, UK: Multilingual Matters *Ltd*.

Department of Basic Education (DBE). 2014. The Ministerial Task Team Report on the National Senior Certificate (NSC): 26 May 2014. Cape Town: Government Printing.

Department of Basic Education (DBE). 2011a. Questions and Answer Booklet for the advocacy of the Implementation of the National Curriculum Statement (NCS) Grade R - 12. Cape Town: Government Printing.

Department of Education. 1997. Language in Education Policy. 14 July 1997. Pretoria: Government Printing.

Du Plessis, C. 2014. Issues of validity and generalisability in the Grade 12 English Home Language examination. *Per Linguam*, 30(2):1-13.

Du Plessis, C. Steyn, S. & Weideman, A. 2013: Towards a construct for assessing high level language ability in Grade 12. Bloemfontein: ICELDA. (unpublished).

Du Plessis, C. Steyn, S. & Weideman, A. 2016. The assessment of home languages in the South African National Senior Certificate examinations – ensuring fairness and increased credibility. *LitNet Akademies*. 28 April 2016.

Du Plessis, C., L., Le, P., L. & Weideman, A. 2017. Test and Content: the use of the Test of Academic Literacy levels (TALL) at tertiary institution in Vietnam. *ResearchGate*, 45(2): 115-131.

Du Plessis, C., Weideman, A., Le, P.L. 2012. Test and content: the use of the Test of Academic Literacy Levels (TALL) at a tertiary institution in Vietnam. *Journal of language teaching*, 45(2) 2011: 115-131.

Du Plessis, C.L. 2017. Developing a theoretical rationale for the attainment of greater equivalence of standard in the Grade 12 Home Language exit-level. University of the Free State. Free State.

Du Plessis, C., Steyn, S. & Weideman, A. 2014. Equalising the South African National Senior Certificate (NSC) Home Language Examinations: A joint project between Umalusi and ICELDA. Preliminary report to the Umalusi Research Forum. Bloemfontein: ICELDA. (unpublished).

Du Plessis, Steyn & Weideman. 2013. Towards a construct for assessing high level language ability in Grade 12 - incomplete

Fulcher, G. & Davidson, F. 2007. Language testing and assessment: *an advanced resource book*. Abingdon: Routledge.

Fulcher, G. & Davidson, F. (eds). 2012. The Routledge handbook of language testing. New York: Routledge

Fulcher, G. 2015. Re-examining language testing: a philosophical and social enquiry. Abingdon, Oxford: Routledge.

Fulscher, G. 2010. Practical language testing. London: Hodder Educational.

Gambier, Y. & Van Doorlaer, L. 2010. Handbook of Translation studies. Amsterdam/Philadelphia: Johan Benjamin Publishing Company.

Gentzler, E. 2001. Contemporary translation theories (2nd ed). Clevedon, Buffalo, Toronto, Sydney: Multilingual matters LTD.

Gierl, M.J. 2000. Construct equivalence on translated achievement tests. *Canadian society for the study of education*.25(4): 280-296.

Gile, D. 1994. Basic concepts and models for interpreter and translation training. Amsterdam: John Benjamins Publishing Company.

Gordad, B. 1990. Theorising feminist discourse/translation, S.Bassnett and A.Lefevere(eds) 87-96.

Green, A. 2014. Exploring language assessment and testing. New York: Routledge.

Green, R. 2013. Statistical analyses for language testers. New York: Palgrave Macmillan.

Habermas, J. 1970. Toward a theory of communicative competence. In Dreitzel, H.P. (Ed.). *Recent sociology* 2. London: Collier-Macmillan. 115-148.

Halliday, M.A.K. 1978. *Language as social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.

Hambleton, R.K. & Patsula, L. 1998. Adapting tests for use in multiple languages and cultures. *Social indicators research*, (45): 153-171.

Hambleton, R.K. & Patsula, L. 2000. Adapting tests for use in multiple languages and culture. *Laboratory of psychometric and evaluative research report*. 459 207: 1-32.

Henning, G. 1987. *A guide to language testing: development, evaluation, research*. Cambridge, MA: Newbury House

Hughes, A. 2003. *Testing for language teachers*. Second edition. Cambridge: Cambridge University Press.

Hymes, D. 1972. On communicative competence. In Pride, J.B. & Holmes, J. (Eds). *Sociolinguistics: Selected readings*. Harmondsworth: Penguin. 269-293.

ICELDA. 2013. Test of advanced language ability: part 1. (unpublished).

ICELDA. 2022. Inter-institutional centre for language development and assessment. <https://icelda.com/> Date accessed: 03 Nov. 2022.

International Test Commission (ITC). 2016. The ITC guidelines for translation and adapting tests (2nd ed). [www,InTestCom,org](http://www.IntTestCom.org). incomplete?

Jansen, D. 1998. Curriculum reform in South Africa: a critical analysis of outcomes-based education. *Cambridge journal of education*. 28(3): 321-331.

Jansen, D. 1999. The school curriculum since apartheid: intersections of politics and policy in the South African transition. *Journal of curriculum studies*, 31(1) 57-67.

Jansen, 2006. Curriculum reform in South Africa: a critical analysis of outcomes-based education. *Cambridge journal of education*, 28(3): 321-331.

Kane, M. 2006. Validation. In Brennan, R. (ed), *Educational Measurement*. 4:17-64.
incomplete

Kane, M.T. 1992. An argument-based approach to validity. *Psychological Bulletin*. 112(3) 527-535.
Kane, M. 2006. Validation. In Brennan, R. (ed), *Educational Measurement*. 4:17-64.

Kanjee, A. & Sayed, Y. 2013. Assessment in post-apartheid South Africa: challenges for improving education quality and learning. *Assessment in education: principles, policy & practice*, 20(4): 442- 469.

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.

Kunnan, A.J. 1997. Connecting fairness with validation in language assessment. Current developments and alternatives in language assessment. Proceedings of LTRC 96. Jyväskylä, Finland: University of Jyväskylä. 85-105.

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: the construction and use of foreign language tests*. New York. McGraw-Hill.

Lambert, J. 1991. In quest of literary maps, in Harald Kittel and Armin Paul Frank (eds) *Interculturality and the historical study of study of literary translations*, Berli:
language ability in Grade 12

Lefevere, A. 1992. *Translating literature: the German tradition from Luther to Rosenzwiweig*, Assen: Van Gorcum.

Linacre, J.M. 2022. Users guide to Winsteps ministep rasch-model computer. <https://www.Winsteps.com/winman/copyright.htm> Date of access: 01 Nov. 2022.

- Lolwana, P. 2013. Where to with matric? *Matric: what is to be done*, 64-71.
- Londe, Z.C. 2005. incomplete
- Lubisi, R.C., & Murphy, R.J.L. 2010. Assessment in South African Schools. *Assessment in education*, 9 (2) 2002: 255-268.
- Mackey, A. & Gass, S.M. 2005. Second language research: methodology and design. New York: Routledge.
- Mathonsi, E. 1988. Black matriculation results: a mechanism of control. Braamfontein: Skotaville.
- Mayburg-Smit, J-M. 2017. The refinement and uses of a test of academic literacy for Grade 10 students. *Journal for language teaching*, 51(1):271-295.
- Menken, K. 2009. No child left behind and its effects on language policy. *Annual review of Applied Linguistics*, (29): 103-117.
- Messick, S. 1980. Test Validity and the Ethics of assessment. *Educational testing service*. 35(11): 1012-1027.
- Messick. 1989. Validity. In R.L.Linn(ed.), *Educational measurement*, (3) 13-103. New York: Macmillan.
- Messick, S. 1990. Validity of Test Interpretation and Use. Princeton: Reports.
- Ministerial Committee Report (MCR), 1998. Investigation into the senior certificate examination. Pretoria: DBE.
- Munday, J. 2012. Introducing translation studies (3rd ed.). London: Routledge.
- Murray, S. 2012. The standards of the National Senior Certificate home language examination: a comparison of South African official languages. Pretoria: Umalusi.

National education policy initiative. 1992. Curriculum: report of the NEP! curriculum research group (Cape Town, Oxford University press/National Education Co-ordinating Committee).

Newmark, P. 1998. A textbook of translation. New York: Prenticehall.

Nida, E., A. 1976. A framework for the analysis and evaluation of theories of translation in Richard Brislin (ed) Translation: applications and research. New York: Gardener Press.

Niranjana, T.1992. Siting translation: history, post-structuralism, and the colonial context, Berkely, CA: university of California Press.

Nord, C. 1997. Translating as a purposeful activity: functionalist approaches explained. Manchester Uk: St Jerome Publishing.

Nzimande, B. 1995. To test or not to test. Paper presented at the congress of Psychometrics of Psychologists and Personnel Practitioners, CSIR conference centre, Pretoria.

Oller, J.W. 1979. Language test at school. Longman Group Limited: London.

Patterson, R. & Weideman, A. 2013. The typicality of academic discourse and its relevance for constructs of academic literacy. *Journal of Language Teaching*, 47(1):107-122.

Pym. A. 2014. Method in translation history. London and New York: Routledge.

Rambiritch, A. 2012. 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. 2013. Validating the test of academic literature for post graduate students (TALPS). *Journal for language teaching*: 47(1) 175-193.

Robinson, D. 1991. The translator's turn. Baltimore: Johns Hopkins University Press

Sebolai, K.W. 2016. The incremental validity of three tests of academic literacy in the context of a South African university of technology. Bloemfontein: UFS. (Thesis-PhD).

Sedibe, K. 2006. Dismantling apartheid education: an overview of change. *Cambridge Journal of education*, 28(3): 269-282.

Shaw, S. & Weir, C.J. 2007. Examining writing: research and practice in assessing second language writing. *Studies in language testing* 26. Cambridge: Cambridge university.

Simon, S. 1996. Gender in translation: Cultural identity and the politics of transmission, London: Routledge.

Snell-Hornby, M. 1991. Translation studies – Art, Science or Utopia? , in Kitty M. van Leuven-Zwart and Ton Naaijken (eds) *Translation studies: the state of the art*, Amsterdam & Atlanta GA: Rodopi, 13-23.

Spolsky, B. 1975. Language testing ~ The problem of validation. *Papers on language testing 1965-1974*) DC:TESOL. 146-153.

Steyn, S. 2013. Towards the development of equal and fair Home Language Assessments:outline of a pilot study. Bloemfontein, ICELDA. (unpublished).

Steyn, S. 2018. A theoretical justification for the design and refinement of a Test of Advanced Language Ability (TALA). Bloemfontein: UFS (Thesis for MA studies)

Swartz, D. 1992. Issues in the reform of the system of certification in South Africa. In E. Unterhalter, H. Wolpe, & T. Botha (eds.), *Education in a future South Africa: Policy issues and transformation* (136-148). Trenton, N: Africa World Press.

Taylor, L. & Weir, C.J. 2005. Multilingualism and Assessment: achieving transparency, assuring quality, sustaining diversity – proceedings of the ALTE Berlin. Studies in language testing 27. Cambridge: Cambridge University.

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.

Toury, G. 1991. What are descriptive studies into translation likely to yield apart from isolated descriptions? In Kitty M. van Leuven-Zwart and Ton Naaikje (eds) Translation studies: *The state of the art, Amsterdam & Atlanta GA*: Rodopi, 179-192.

Trumplemann, M.H. 1991. The Joint Matriculation Board – seventy-five years achievement in perspective. Cape Town. National Book Printers.

Umalusi. 2007. Making educational judgements: reflections on standards of intended and examined curricula. Pretoria: Umalusi.

Umalusi. 2011. Understanding statistical moderation. Pretoria. Umatlusi.

Van der Slik, F. & Weideman, A. 2005. The refinement of a test of academic literacy. *Per Linguam*, 21(2): 23-35.

Van der Slik, F. & Weideman, A. 2008. The stability of test design: measuring differences in performance across several administrations of academic literacy test. *Acta Academia*, 40(1):161-182.

Van der Slik, F., & Weideman, A. 2010. Examining bias in a test of academic literacy: does the Test of Academic Literacy Level (TALL) treat students from English and African language backgrounds differently? *Journal of language teaching*. 44(2) 2010: 107-150.

Van der Walt, J.L. & Steyn, F. 2008. The validation of language test. *Stellenbosch papers in linguistics*. 2008 (38): 191-204.

Van der Walt, J.L. & Steyn, H.S. 2007. Pragmatic validation of a test of academic literacy at tertiary level. *Ensovoort*, 11(1):138-153.

Van Dyk, T. & Weideman, A. 2004a. Finding the right measure: From blueprint to specification to item type. *Journal for language teaching*, 38(1): 15-24.

Van Dyk, T. & Weideman, A. 2004b. Switching constructs: On the selection of an appropriate blueprint for academic literacy assessment. *SAALT Journal for language teaching*, 38(1): 1-13.

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. 2011. Onderzoek na die impak van akademiese geletterheidsintervensie op eerstejaarstudente taalvermoe. *LiNet Akademies*, 8(3): 487-507.

Van Dyk, T., Van Rensburg, A., & Marais, F. 2011. Levelling the playing field: an investigation into the translation of academic literacy tests. *Journal for Language Teaching*. 45(1):153 -163.

Vermeer, H. 1989. Skopos and commission in translational action. Translated by A. Chesterman. In: Venuti, L. (ed). 2004. The translation studies reader. London: Routledge. 221 - 232.

Vermeer, H. 2000. Skopos and transmission in translational action, in Lawrence Venuti (ed) The translation studies reader, London and New York: Routledge.

Vinay, J-P, & Darbelnet, J. 1995. Comparative stylistic of French and English: a methodology for translation, translated and edited by Juan Sager, Marie-Jo Hamel, Amsterdam and Philadelphia: John Benjamins.

Venuti, L. 1995. The translator's invisibility: a history of translation. London: Routledge.

Von Flotow, L. 2016. *A companion of translation studies*. Ottawa: University of Ottawa Press.

Ward-Cox, M.W. 2020. Validation of a rating scale for distance education university student essay in a literature-based module. Pretoria: Unisa.

Weideman, A. 2008. Assessing academic literacy: A task-based approach. *Language matters*, 37(1): 81-101.

Weideman, A. 2009. Constitutive and regulative conditions for the assessment of academic literacy. *Southern African linguistics and applied language studies*, 27(3): 235-251.

Weideman, A. 2011. Academic literacy tests: design, development, piloting and refinement. *Journal of language teaching*, 45(2): 100-113.

Weideman, A. 2012. Validation and validity beyond Messick. *Per-Linguam*, 28(2): 1-14.

Weideman, A. 2016. The refinement of the idea of consequential validity within an alternative framework for responsible test design. Routeledge.
<https://www.taylorfrancis.com/chapters/edit/10.4324/9781315517377-12/refinement-idea-consequential-validity-within-alternative-framework-responsible-test-design-albert-weideman>. Date of access: 30 Mar 2023.

Weideman, A. 2017. The initial validation of a test of emergent literacy. *Perlinguam*, 33(1): 25-53.

Weideman, A. 2020. Complementary evidence in the early-stage validation of language tests: classical test theory and rasch analysis. *Perlinguam*: 36(2):57-75

Weir, C.J. 2005a. *Language testing and validation: An evidence-based approach*. New York: Palgrave Macmillan.

Weir, C.J. 2005b. Limitations of the Common European Framework for developing comparable examinations and tests. *Language Testing*, (22):281-300.

William, N. 2011. Research methods: the basics. London: Routledge.

Zohre Owji, M.A. 2013. Translation strategies: a review and comparison of theories. *Translation journal*. 17(1) 1-3.

APPENDICES

APPENDIX A1: TIAPLUS REPORT ON THE RESULTS OF THE ANALYSIS OF THE 187 ITEMS OF THE T-SES-TALA PART 1 AND 2

iaPlus® Test and Item Analysis Build 314
Cito, Measurement and Research Department. Arnhem, the Netherlands. © 2013.

Population : NWU students
Test : TALA (Sesotho translation)
Date : Tuesday, 28 September 2021
Time : 10:10
Data file : C:\Albert Documents\ICELDA\ice02623_TALA_Sesotho_187_item.txt
Database : C:\Cito\TiaPlus\TiaPlus2000.mdb
Missing handling : Missing as Zero

Persons: All persons

Items: All items

Test and Item Analysis

Item Label	Item nr.	Weight	Key	P- and A- values					§ # Mis- sing					Weighted									
				A	B	C	D	E	F	O/D	Max	Mean	P	Sd	RSK	[Lo Rit Hi]	Rir	AR					
	1	1	C	10	7	59*	8	13				3	14	1	0,59	59	0,49	0,49	6	15	23	8	40
	2	1	A	28*	17	13	12	25				4	19	1	0,28	28	0,45	0,45	-1	9	18	3	41
	3	1	E	18	34	12	18	14*				4	20	1	0,14	14	0,35	0,35	5	14	22	9	40
	4	1	B	17	28*	6	25	20				4	23	1	0,28	28	0,45	0,45	-2	6	14	0	41
	5	1	D	26	12	8	34*	15				5	27	1	0,34	34	0,47	0,47	5	13	22	7	40
	6	1	D	5	25	8	53*	6				3	16	1	0,53	53	0,50	0,50	10	19	27	12	40
	7	1	E	26	18	27	13	13*				3	16	1	0,13	13	0,34	0,34	-1	7	15	3	41
	8	1	C	25	20	20*	11	20				4	22	1	0,20	20	0,40	0,40	-2	6	15	1	41
	9	1	B	26	14*	21	11	24				4	20	1	0,14	14	0,35	0,35	6	14	23	10	40
	10	1	A	17*	23	24	8	24				4	20	1	0,17	17	0,37	0,37	-11	-1	8	-6	41
	11	1	C	50	11	28*	9	0				3	14	1	0,28	28	0,45	0,45	-1	7	16	2	41
	12	1	D	33	20	3	43*	0				2	10	1	0,43	43	0,49	0,49	-10	-1	7	-8	41
	13	1	B	19	36*	24	18	0				3	14	1	0,36	36	0,48	0,48	1	10	18	3	41
	14	1	A	34*	10	11	40	1				3	14	1	0,34	34	0,48	0,48	1	9	17	3	41
	15	1	D	11	5	33	47*	1				3	13	1	0,47	47	0,50	0,50	10	18	26	12	40
	16	1	C	20	24	15*	40	0				1	5	1	0,15	15	0,36	0,36	-14	-6	2	-11	41
	17	1	A	61*	9	13	15	1				1	4	1	0,61	61	0,49	0,49	3	12	20	5	40
	18	1	C	19	27	32*	21	0				2	10	1	0,32	32	0,46	0,46	-7	2	11	-4	41
	19	1	A	14*	51	16	18	0				1	3	1	0,14	14	0,35	0,35	-14	-6	2	-11	41
	20	1	C	26	28	30*	11	0				5	24	1	0,30	30	0,46	0,46	-3	5	14	-1	41
	21	1	C	27	26	34*	10	0				3	17	1	0,34	34	0,47	0,47	-1	8	17	2	41
	22	1	C	22	11	46*	16	0				5	26	1	0,46	46	0,50	0,50	0	9	17	2	41
	23	1	D	23	25	18	31*	0				2	10	1	0,31	31	0,46	0,46	3	11	20	5	40
	24	1	D	15	15	43	26*	0				2	9	1	0,26	26	0,44	0,44	-4	5	13	-1	41
	25	1	C	17	35	27*	18	1				2	10	1	0,27	27	0,45	0,45	-5	4	13	-2	41
	26	1	B	16	39*	23	19	0				2	11	1	0,39	39	0,49	0,49	-7	1	10	-5	41
	27	1	A	29*	31	30	9	0				1	4	1	0,29	29	0,45	0,45	-3	6	14	0	41
	28	1	C	32	23	25*	18	0				2	10	1	0,25	25	0,43	0,43	2	11	20	5	40
	29	1	C	18	18	52*	10	0				2	8	1	0,52	52	0,50	0,50	18	26	34	20	39
	30	1	B	26	42*	17	10	0				4	21	1	0,42	42	0,49	0,49	11	19	27	13	40
	31	1	C	25	17	30*	26	0				1	5	1	0,30	30	0,46	0,46	5	13	21	8	40
	32	1	B	37	42*	10	9	0				2	10	1	0,42	42	0,49	0,49	5	13	22	7	40
	33	1	C	20	27	25*	26	0				2	12	1	0,25	25	0,43	0,43	3	11	19	5	40
	34	1	A	46*	26	14	11	1				2	10	1	0,46	46	0,50	0,50	1	10	18	4	41
	35	1	D	26	23	21	26*	1				3	15	1	0,26	26	0,44	0,44	5	13	21	8	40
	36	1	D	35	33	11	17*	1				2	9	1	0,17	17	0,38	0,38	-4	3	10	-2	41
	37	1	B	28	21*	36	14	0				1	5	1	0,21	21	0,40	0,40	-7	2	10	-4	41
	38	1	C	18	28	40*	11	1				3	14	1	0,40	40	0,49	0,49	9	18	26	11	40
	39	1	A	33*	26	20	18	0				2	12	1	0,33	33	0,47	0,47	9	17	25	11	40
	40	1	A	38*	21	18	20	1				2	11	1	0,38	38	0,49	0,49	5	14	22	7	40
	41	1	C	26	34	27*	11	0				2	8	1	0,27	27	0,44	0,44	0	8	16	2	41
	42	1	A	24*	21	18	34	1				2	8	1	0,24	24	0,43	0,43	-15	-6	3	-11	42
	43	1	A	30*	29	29	10	0				2	11	1	0,30	30	0,46	0,46	1	9	17	3	41
	44	1	B	16	44*	20	18	1				2	8	1	0,44	44	0,50	0,50	6	15	23	8	40
	45	1	D	16	18	21	41*	1				3	13	1	0,41	41	0,49	0,49	1	9	18	3	41
	46	1	D	36	19	21	23*	0				1	5	1	0,23	23	0,42	0,42	5	13	21	8	40
	47	1	B	21	39*	22	16	0				1	7	1	0,39	39	0,49	0,49	2	11	19	4	40
	48	1	A	25*	32	26	16	0				2	10	1	0,25	25	0,43	0,43	6	14	22	9	40
	49	1	C	31	27	22*	18	0				2	9	1	0,22	22	0,41	0,41	-5	4	12	-1	41
	50	1	C	12	19	53*	14	1				2	8	1	0,53	53	0,50	0,50	-1	8	16	1	41
	51	1	B	22	32*	22	23	1				2	8	1	0,32	32	0,47	0,47	-2	6	15	0	41
	52	1	D	25	27	31	14*	0				2	11	1	0,14	14	0,35	0,35	-4	4	12	-1	41
	53	1	D	25	28	25	19*	0				3	13	1	0,19	19	0,40	0,40	-6	3	12	-2	41
	54	1	D	27	26	25	19*	0				4	18	1	0,19	19	0,39	0,39	-12	-3	6	-8	41
	55	1	D	25	23	21	25*	1				4	23	1	0,25	25	0,44	0,44	-3	5	13	0	41
	56	1	D	26	24	18	29*	1				2	11	1	0,29	29	0,45	0,45	-3	5	13	-1	41
	57	1	C	25	31	25*	16	0				2	12	1	0,25	25	0,44	0,44	-3	5	14	0	41
	58	1	A	22*	30	29	15	1				3	17	1	0,22	22	0,41	0,41	-5	4	12	-2	41
	59	1	A	18*	35	26	17	0				4	19	1	0,18	18	0,38	0,38	5	14	23	9	40
	60	1	C	26	26	19*	23	1				4	23	1	0,19	19	0,39	0,39	0	8	15	3	41
	61	1	C	30	29	22*	16	0				3	14	1	0,22	22	0,41	0,41	-5	3	11	-2	41

62	1	A	24*	30	27	15	1	3	13		1	0,24	24	0,43	0,43	-3	5	14	0	41
63	1	D	27	26	25	17*	1	4	18		1	0,17	17	0,38	0,38	-2	6	15	1	41
64	1	D	21	28	24	24*	1	3	13		1	0,24	24	0,43	0,43	2	11	19	5	40
65	1	C	21	27	26*	20	1	4	23		1	0,26	26	0,44	0,44	-2	6	15	1	41
66	1	A	33*	25	21	17	0	3	17		1	0,33	33	0,47	0,47	6	14	22	8	40
67	1	D	21	28	29	19*	0	3	16		1	0,19	19	0,39	0,39	-3	5	13	0	41
68	1	B	21	28*	27	19	1	4	20		1	0,28	28	0,45	0,45	4	13	21	7	40
69	1	C	21	28	27*	18	0	4	22		1	0,27	27	0,45	0,45	1	10	18	4	41
70	1	B	21	29*	25	19	1	6	29		1	0,29	29	0,45	0,45	0	9	18	3	41
71	1	A	29*	23	22	20	1	5	25		1	0,29	29	0,45	0,45	5	14	22	8	40
72	1	B	21	35*	24	15	0	5	25		1	0,35	35	0,48	0,48	7	16	24	10	40
73	1	B	22	28*	29	14	1	5	28		1	0,28	28	0,45	0,45	-7	2	10	-4	41
74	1	C	20	24	31*	21	0	4	20		1	0,31	31	0,46	0,46	-3	6	14	0	41
75	1	D	24	25	25	21*	1	4	21		1	0,21	21	0,41	0,41	-4	4	12	-1	41
76	1	A	39*	25	18	14	0	3	15		1	0,39	39	0,49	0,49	-1	8	16	1	41
77	1	A	21*	29	29	18	0	3	17		1	0,21	21	0,41	0,41	4	13	22	8	40
78	1	B	30	29*	23	15	0	4	18		1	0,29	29	0,45	0,45	0	9	17	3	41
79	1	B	21	26*	28	20	0	5	24		1	0,26	26	0,44	0,44	-4	4	12	-2	41
80	1	C	29	27	20*	16	1	7	34		1	0,20	20	0,40	0,40	-9	-1	7	-6	41
81	1	A	22*	32	27	16	0	4	21		1	0,22	22	0,41	0,41	-4	5	14	0	41
82	1	C	22	31	25*	19	0	3	15		1	0,25	25	0,43	0,43	2	10	18	4	41
83	1	B	24	38*	22	13	0	4	18		1	0,38	38	0,48	0,48	6	15	23	8	40
84	1	C	26	27	21*	24	0	3	15		1	0,21	21	0,41	0,41	5	13	21	8	40
85	1	A	14*	39	28	16	0	4	18		1	0,14	14	0,34	0,34	-3	6	14	1	41
86	1	A	34*	22	20	19	0	4	21		1	0,34	34	0,48	0,48	10	18	25	12	40
87	1	A	24*	24	28	19	0	5	28		1	0,24	24	0,43	0,43	-2	6	14	1	41
88	1	D	35	18	20	20*	0	7	38		1	0,20	20	0,40	0,40	0	8	16	3	41
89	1	B	27	30*	21	19	0	3	17		1	0,30	30	0,46	0,46	-1	7	16	2	41
90	1	B	15	37*	27	18	0	4	19		1	0,37	37	0,48	0,48	8	16	24	10	40
91	1	C	21	30	29*	15	0	6	31		1	0,29	29	0,45	0,45	2	11	19	5	40
92	1	C	19	27	29*	18	0	6	33		1	0,29	29	0,45	0,45	2	10	18	4	41
93	1	A	26*	26	18	25	0	6	29		1	0,26	26	0,44	0,44	6	14	22	8	40
94	1	C	30	26	22*	19	0	3	15		1	0,22	22	0,42	0,42	1	10	18	4	41
95	1	C	18	29	33*	17	0	2	12		1	0,33	33	0,47	0,47	-3	6	15	0	41
96	1	B	17	32*	30	15	0	6	29		1	0,32	32	0,47	0,47	-4	4	12	-2	41
97	1	B	19	32*	26	18	0	5	24		1	0,32	32	0,46	0,46	1	10	18	4	41
98	1	C	24	24	25*	22	0	5	26		1	0,25	25	0,43	0,43	2	10	19	5	40
99	1	C	32	29	22*	15	0	4	18		1	0,22	22	0,41	0,41	0	8	16	3	41
100	1	D	19	31	33	13*	0	4	22		1	0,13	13	0,33	0,33	-9	0	9	-4	41
101	1	D	17	35	27	16*	0	4	23		1	0,16	16	0,37	0,37	5	13	21	8	40
102	1	B	22	29*	25	21	0	4	19		1	0,29	29	0,45	0,45	1	9	17	3	41
103	1	D	24	26	26	16*	0	7	35		1	0,16	16	0,37	0,37	3	11	18	6	40
104	1	C	32	27	23*	14	0	4	20		1	0,23	23	0,42	0,42	-2	6	14	0	41
105	1	A	20*	35	29	12	0	4	23		1	0,20	20	0,40	0,40	1	9	16	4	41
106	1	C	21	33	25*	15	0	6	30		1	0,25	25	0,43	0,43	-2	6	14	0	41
107	1	C	21	29	28*	18	0	4	20		1	0,28	28	0,45	0,45	-10	-1	7	-7	41
108	1	B	23	22*	28	22	0	5	25		1	0,22	22	0,42	0,42	3	11	19	6	40
109	1	B	32	30*	22	13	0	3	16		1	0,30	30	0,46	0,46	3	12	20	6	40
110	1	C	20	35	28*	14	0	4	18		1	0,28	28	0,45	0,45	0	8	16	2	41
111	1	C	20	33	30*	14	0	4	18		1	0,30	30	0,46	0,46	-5	3	11	-3	41
112	1	D	21	30	26	19*	0	3	17		1	0,19	19	0,39	0,39	-4	4	12	-1	41
113	1	A	27*	25	26	18	0	4	20		1	0,27	27	0,44	0,44	1	10	19	4	40
114	1	A	30*	29	22	17	0	2	11		1	0,30	30	0,46	0,46	0	8	16	2	41
115	1	A	18*	36	27	15	0	4	21		1	0,18	18	0,38	0,38	6	14	23	10	40
116	1	C	16	31	35*	15	0	4	22		1	0,35	35	0,48	0,48	8	16	24	10	40
117	1	A	23*	30	26	16	0	4	22		1	0,23	23	0,42	0,42	4	12	20	7	40
118	1	B	26	27*	24	18	0	5	25		1	0,27	27	0,44	0,44	-4	4	12	-2	41
119	1	B	36	29*	16	15	0	4	19		1	0,29	29	0,45	0,45	10	18	26	12	40
120	1	A	18*	35	30	15	0	3	15		1	0,18	18	0,38	0,38	7	14	22	9	40
121	1	D	20	31	29	17*	0	4	19		1	0,17	17	0,38	0,38	4	13	22	8	40
122	1	C	20	31	29*	16	0	5	24		1	0,29	29	0,45	0,45	2	11	19	5	40
123	1	B	21	25*	27	22	0	5	25		1	0,25	25	0,44	0,44	3	11	19	5	40
124	1	A	31*	26	21	19	0	3	14		1	0,31	31	0,46	0,46	6	14	22	8	40
125	1	D	16	32	30	18*	0	4	21		1	0,18	18	0,38	0,38	6	15	23	10	40
126	1	B	17	32*	30	18	0	3	16		1	0,32	32	0,47	0,47	8	16	24	10	40
127	1	B	19	31*	30	15	0	4	23		1	0,31	31	0,46	0,46	-4	4	13	-2	41
128	1	A	28*	25	26	15	0	5	28		1	0,28	28	0,45	0,45	14	22	30	17	40
129	1	B	31	32*	18	14	0	4	20		1	0,32	32	0,47	0,47	7	15	23	9	40
130	1	B	17	35*	27	16	0	5	26		1	0,35	35	0,48	0,48	-2	6	15	0	41
131	1	A	21*	28	31	15	0	4	23		1	0,21	21	0,41	0,41	7	15	23	10	40
132	1	A	20*	31	26	18	0	6	31		1	0,20	20	0,40	0,40	10	18	27	13	40
133	1	C	27	25	23*	20	0	5	27		1	0,23	23	0,42	0,42	-4	4	13	-1	41
134	1	B	37	27*	20	13	0	3	16		1	0,27	27	0,45	0,45	8	16	25	11	40
135	1	B	16	36*	28	16	0	4	22		1	0,36	36	0,48	0,48	-1	8	16	2	41
136	1	A	20*	34	30	13	0	4	18		1	0,20	20	0,40	0,40	-6	3	11	-2	41
137	1	D	23	24	28	20*	0	5	26		1	0,20	20	0,40	0,40	8	16	24	11	40
138	1	B	26	24*	28	15	0	6	32		1	0,24	24	0,43	0,43	3	12	20	6	40
139	1	B	32	30*	18	15	0	4	22		1	0,30	30	0,46	0,46	16	24	31	18	40
140	1	B	18	35*	28	14	0	5	24		1	0,35	35	0,48	0,48	10	18	26	12	40
141	1	C	18	36	25*	15	0	5	28		1	0,25	25	0,44	0,44	5	14	22	8	40
142	1	B	22	29*	29	14	0	5	24		1	0,29	29	0,46	0,46	13	21	29	15	40
143	1	C	22	28	25*	19	0	6	32		1	0,25	25	0,43	0,43	4	12	20	6	40
144	1	C	25	27	26*	18	0	4	20		1	0,26	26	0,44	0,44	5	13	21	8	40
145	1	A	18*	35	29															

155	1	D	17	31	29	15*	0	8	40		1	0,15	15	0,36	0,36	-6	2	10	-3	41
156	1	A	18*	31	27	16	0	7	38		1	0,18	18	0,38	0,38	-4	3	11	-2	41
157	1	A	21*	27	27	17	0	8	40		1	0,21	21	0,41	0,41	0	8	17	3	41
158	1	A	25*	25	23	20	0	7	38		1	0,25	25	0,43	0,43	-1	7	15	1	41
159	1	A	36*	25	21	13	0	6	30		1	0,36	36	0,48	0,48	-4	4	13	-2	41
160	1	D	16	32	32	14*	0	7	36		1	0,14	14	0,35	0,35	0	8	15	3	41
161	1	C	22	31	29*	12	0	5	28		1	0,29	29	0,46	0,46	8	16	24	10	40
162	1	C	20	28	27*	18	0	8	39		1	0,27	27	0,44	0,44	4	12	20	6	40
163	1	B	25	24*	25	17	0	9	44		1	0,24	24	0,43	0,43	5	13	21	7	40
164	1	B	31	27*	19	14	0	8	41		1	0,27	27	0,44	0,44	8	17	25	11	40
165	1	C	17	29	33*	14	0	7	35		1	0,33	33	0,47	0,47	9	17	24	11	40
166	1	D	17	30	31	14*	0	8	41		1	0,14	14	0,35	0,35	-1	8	16	3	41
167	1	D	21	29	27	16*	0	7	37		1	0,16	16	0,37	0,37	-6	3	11	-2	41
168	1	A	29*	22	24	18	0	8	42		1	0,29	29	0,45	0,45	4	12	20	6	40
169	1	B	34	26*	21	11	0	7	35		1	0,26	26	0,44	0,44	-1	8	17	2	41
170	1	B	18	27*	33	13	0	9	47		1	0,27	27	0,44	0,44	0	9	17	3	41
171	1	C	18	30	26*	18	0	8	42		1	0,26	26	0,44	0,44	3	11	19	6	40
172	1	B	17	32*	28	17	0	7	36		1	0,32	32	0,47	0,47	8	16	24	10	40
173	1	C	24	25	25*	17	0	9	45		1	0,25	25	0,43	0,43	6	14	22	8	40
174	1	A	32*	26	18	16	0	8	40		1	0,32	32	0,47	0,47	5	13	21	7	40
175	1	C	11	36	31*	15	0	7	38		1	0,31	31	0,46	0,46	8	16	24	10	40
176	1	C	20	27	29*	15	0	8	40		1	0,29	29	0,46	0,46	0	8	17	3	41
177	1	A	20*	33	24	15	0	7	38		1	0,20	20	0,40	0,40	-3	5	13	0	41
178	1	A	24*	25	23	20	0	8	40		1	0,24	24	0,43	0,43	4	13	21	7	40
179	1	D	26	26	22	18*	0	8	42		1	0,18	18	0,38	0,38	1	9	16	4	41
180	1	D	16	34	30	13*	0	7	38		1	0,13	13	0,34	0,34	-10	-2	7	-6	41
181	1	B	17	32*	30	13	0	7	37		1	0,32	32	0,47	0,47	2	10	18	4	41
182	1	B	19	31*	25	18	0	7	36		1	0,31	31	0,46	0,46	0	9	17	3	41
183	1	C	30	24	24*	15	0	7	37		1	0,24	24	0,43	0,43	-6	3	11	-3	41
184	1	A	32*	22	15	15	0	17	88		1	0,32	32	0,47	0,47	2	10	19	4	40
185	1	D	15	25	29	13*	0	18	91		1	0,13	13	0,33	0,33	-6	3	12	-1	41
186	1	C	20	26	21*	16	0	18	90		1	0,21	21	0,41	0,41	-5	4	13	-1	41
187	1	B	19	20*	17	19	0	25	129		1	0,20	20	0,40	0,40	-2	7	16	2	41

SubGroup number			: 0					SubTest number			: 0									
Number of persons in test			: 514					Number of selected items			: 187									
Minimum test score			: 0					Maximum test score			: 187									
Average test score			: 50,53					Standard deviation			: 7,72									
Average P-value			: 27,02					Std. Error of Measurement			: 5,94									
Average Rit			: 0,09																	
Coefficient Alpha			: 0,41					SE Coeff. Alpha			: 0,04									
GLB			: 0,00					Asymptotic GLB coef			: 0,00									
Items used in GLB proc			: 187																	
Cut-off score			: 60,5					Percentage failing			: 91,05									

Misclassifications:																				
Alpha based										GLB based										

-Rxx' case	Percentage	: 14,4						Percentage	: 17,7											
	Number	: 74						Number	: 91											
-Rxt case	Percentage	: 11,5						Percentage	: 17,7											
	Number	: 59						Number	: 91											

90% Confidence limits for Coefficient Alpha: (0,34 =< 0,41 =< 0,47)																				

APPENDIX A2: TIAPLUS REPORT ON THE RESULTS OF THE ANALYSIS OF THE 187 ITEMS OF THE NC-SES-TALA PART 1 AND 2

TiaPlus® Test and Item Analysis Build 314
Cito, Measurement and Research Department. Arnhem, the Netherlands. © 2013.

Population : NWU students
Test : TALA (new Sesotho version)
Date : Thursday, 30 September 2021
Time : 12:42
Data file : C:\Albert Documents\ICELDA\ice02636_TALA_Sesotho2.txt
Database : C:\Cito\TiaPlus\TiaPlus2000.mdb
Missing handling : Missing as Zero

Persons: All persons

Items: All items

Test and Item Analysis

Item Label	Item nr.	Weight	Key	P- and A- values					% #					Weighted									
				A	B	C	D	E	F	O/D	Mis- sing	Max	Mean	P	Sd	RSK	[Lo	Rit	Hi]	Rir	AR		
	1	1	C	23	22	42*	9	4			1	5	1	0,42	42	0,49	0,49	20	28	36	25	82	
	2	1	E	18	25	25	22	7*			2	12	1	0,07	7	0,25	0,25	0	9	17	7	82	
	3	1	A	19*	26	22	17	12			4	19	1	0,19	19	0,40	0,40	-7	0	7	-3	82	
	4	1	D	20	27	13	23*	13			4	19	1	0,23	23	0,42	0,42	-5	3	11	0	82	
	5	1	B	20	21*	18	17	20			4	20	1	0,21	21	0,41	0,41	-4	5	14	2	82	
	6	1	D	20	29	18	18*	13			2	10	1	0,18	18	0,39	0,39	0	9	18	6	82	
	7	1	E	16	26	28	17	10*			3	17	1	0,10	10	0,30	0,30	17	27	37	25	82	
	8	1	A	24*	24	24	17	7			3	17	1	0,24	24	0,43	0,43	0	8	17	5	82	
	9	1	C	24	20	24*	20	8			3	16	1	0,24	24	0,43	0,43	-3	6	14	3	82	
	10	1	B	21	22*	24	16	12			5	27	1	0,22	22	0,42	0,42	-14	-5	3	-8	82	
	11	1	C	23	23	41*	10	1			1	6	1	0,41	41	0,49	0,49	26	33	40	30	82	
	12	1	D	14	20	25	38*	2			1	7	1	0,38	38	0,49	0,49	27	35	43	32	82	
	13	1	A	24*	29	26	20	1			1	7	1	0,24	24	0,42	0,42	1	10	20	7	82	
	14	1	A	32*	25	24	15	2			2	10	1	0,32	32	0,47	0,47	22	30	38	27	82	
	15	1	C	16	18	44*	17	3			3	14	1	0,44	44	0,50	0,50	34	41	48	38	81	
	16	1	C	20	16	45*	17	1			1	4	1	0,45	45	0,50	0,50	30	37	44	34	82	
	17	1	D	21	21	22	34*	0			2	10	1	0,34	34	0,47	0,47	34	42	49	39	81	
	18	1	A	26*	25	33	12	1			3	13	1	0,26	26	0,44	0,44	10	19	28	16	82	
	19	1	D	19	21	20	36*	2			2	11	1	0,36	36	0,48	0,48	34	41	48	38	81	
	20	1	C	20	21	38*	14	3			3	17	1	0,38	38	0,49	0,49	36	43	49	40	81	
	21	1	C	37	23	24*	12	2			2	12	1	0,24	24	0,43	0,43	10	19	27	16	82	
	22	1	B	15	43*	27	12	1			2	12	1	0,43	43	0,50	0,50	23	31	39	28	82	
	23	1	D	19	31	19	26*	2			3	13	1	0,26	26	0,44	0,44	13	23	32	20	82	
	24	1	D	19	25	24	29*	2			2	11	1	0,29	29	0,45	0,45	24	33	42	30	82	
	25	1	C	18	21	39*	16	2			3	18	1	0,39	39	0,49	0,49	33	41	47	38	81	
	26	1	C	27	22	35*	14	1			1	7	1	0,35	35	0,48	0,48	49	55	61	53	81	
	27	1	B	30	31*	25	9	1			3	14	1	0,31	31	0,46	0,46	-5	4	12	0	82	
	28	1	C	16	23	45*	13	2			2	8	1	0,45	45	0,50	0,50	35	42	48	39	81	
	29	1	C	20	23	34*	18	2			2	12	1	0,34	34	0,48	0,48	32	39	47	36	81	
	30	1	C	26	22	32*	15	2			3	13	1	0,32	32	0,47	0,47	19	28	36	24	82	
	31	1	D	28	22	20	26*	2			2	8	1	0,26	26	0,44	0,44	28	37	46	34	82	
	32	1	B	13	44*	27	13	0			2	11	1	0,44	44	0,50	0,50	27	35	42	31	82	
	33	1	C	19	31	30*	17	0			2	12	1	0,30	30	0,46	0,46	9	18	26	15	82	
	34	1	C	17	22	40*	16	1			3	16	1	0,40	40	0,49	0,49	21	29	37	26	82	
	35	1	B	26	38*	18	13	3			3	14	1	0,38	38	0,49	0,49	33	40	47	37	81	
	36	1	D	35	19	17	27*	1			2	8	1	0,27	27	0,44	0,44	16	25	34	22	82	
	37	1	C	28	29	27*	15	0			2	8	1	0,27	27	0,44	0,44	1	11	20	7	82	
	38	1	C	15	29	42*	12	1			2	11	1	0,42	42	0,49	0,49	25	32	39	29	82	
	39	1	C	25	24	32*	16	1			3	14	1	0,32	32	0,47	0,47	11	20	28	17	82	
	40	1	C	27	21	29*	18	3			3	15	1	0,29	29	0,45	0,45	15	24	32	21	82	
	41	1	C	30	19	35*	13	1			2	8	1	0,35	35	0,48	0,48	31	38	45	35	82	
	42	1	B	20	40*	26	11	1			2	8	1	0,40	40	0,49	0,49	2	11	19	7	82	
	43	1	B	25	37*	21	14	1			2	12	1	0,37	37	0,48	0,48	13	21	30	18	82	
	44	1	B	18	33*	26	18	1			4	21	1	0,33	33	0,47	0,47	11	20	29	17	82	
	45	1	C	18	24	36*	16	2			4	22	1	0,36	36	0,48	0,48	33	40	47	38	81	
	46	1	C	27	24	33*	12	1			3	14	1	0,33	33	0,47	0,47	29	37	44	34	82	
	47	1	B	18	34*	30	14	1			2	12	1	0,34	34	0,48	0,48	2	11	19	8	82	
	48	1	C	18	30	34*	13	2			3	14	1	0,34	34	0,47	0,47	1	9	18	6	82	
	49	1	D	20	33	21	23*	1			3	16	1	0,23	23	0,42	0,42	13	23	32	20	82	
	50	1	C	24	21	31*	18	2			3	18	1	0,31	31	0,46	0,46	24	32	40	29	82	
	51	1	A	46*	22	16	12	2			3	14	1	0,46	46	0,50	0,50	15	23	30	19	82	
	52	1	D	10	31	28	27*	1			3	15	1	0,27	27	0,44	0,44	31	39	47	37	82	
	53	1	C	15	30	36*	16	1			3	16	1	0,36	36	0,48	0,48	13	22	30	18	82	
	54	1	C	14	25	33*	23	1			3	17	1	0,33	33	0,47	0,47	22	31	39	28	82	
	55	1	B	22	35*	20	17	2			4	20	1	0,35	35	0,48	0,48	24	32	40	29	82	
	56	1	A	39*	25	18	13	2			3	16	1	0,39	39	0,49	0,49	18	26	34	23	82	
	57	1	D	15	34	27	20*	2			2	11	1	0,20	20	0,40	0,40	-2	8	17	5	82	
	58	1	A	26*	26	32	12	1			3	18	1	0,26	26	0,44	0,44	17	26	35	23	82	
	59	1	A	27*	28	22	18	2			3	16	1	0,27	27	0,45	0,45	19	28	37	25	82	
	60	1	D	21	26	22	25*	1			4	23	1	0,25	25	0,43	0,43	17	25	34	22	82	
	61	1	A	37*	25	20	14	1			3	14	1	0,37	37	0,48	0,48	1	10	19	7	82	
	62	1	C	14	34	31*	15	1			4	23	1	0,31	31	0,46	0,46	-4	5	14	2	82	
	63	1	A	17*	33	31	15	1			3	15	1	0,17	17	0,38	0,38	-5	3	12	1	82	
	64	1	A	21*	29	23	20	2			4	22	1	0,21	21	0,41	0,41	-1	7	16	4	82	
	65	1	C	24	21	30*	18	1			6	29	1	0,30	30	0,46	0,46	11	20	29	17	82	
	66	1	B	28	29*	23	15	1			4	20	1	0,29	29	0,45	0,45	12	21	30	18	82	
	67	1	C	13	28	39*	16	0			3	16	1	0,39	39	0,49							

68	1	B	21	32*	28	15	1	3	18		1	0,32	32	0,46	0,46	5	14	23	11	82
69	1	D	18	32	27	19*	1	3	14		1	0,19	19	0,39	0,39	-8	0	8	-3	82
70	1	B	25	31*	23	16	2	4	19		1	0,31	31	0,46	0,46	2	11	19	7	82
71	1	B	30	25*	21	18	1	5	24		1	0,25	25	0,44	0,44	-5	3	11	0	82
72	1	C	13	31	37*	13	1	5	24		1	0,37	37	0,48	0,48	16	25	33	21	82
73	1	A	24*	30	26	15	1	4	23		1	0,24	24	0,43	0,43	17	25	34	22	82
74	1	A	22*	24	30	19	1	4	23		1	0,22	22	0,41	0,41	11	20	29	17	82
75	1	D	27	23	19	24*	2	5	28		1	0,24	24	0,43	0,43	24	33	41	30	82
76	1	C	34	23	26*	13	1	3	13		1	0,26	26	0,44	0,44	23	32	40	29	82
77	1	B	13	39*	27	15	1	4	23		1	0,39	39	0,49	0,49	3	12	21	9	82
78	1	B	17	37*	29	13	0	4	19		1	0,37	37	0,48	0,48	6	14	23	11	82
79	1	D	21	26	24	25*	1	3	14		1	0,25	25	0,43	0,43	3	13	23	10	82
80	1	B	30	28*	24	12	2	4	22		1	0,28	28	0,45	0,45	0	9	17	6	82
81	1	C	31	29	21*	14	1	4	20		1	0,21	21	0,41	0,41	0	9	19	6	82
82	1	A	31*	25	29	10	1	4	21		1	0,31	31	0,46	0,46	29	37	45	34	82
83	1	B	25	35*	24	12	1	3	17		1	0,35	35	0,48	0,48	9	18	26	14	82
84	1	B	27	24*	32	14	0	2	11		1	0,24	24	0,43	0,43	0	9	18	6	82
85	1	A	10*	38	32	15	0	4	22		1	0,10	10	0,30	0,30	-8	1	10	-2	82
86	1	B	11	24*	45	16	0	4	19		1	0,24	24	0,43	0,43	-9	-1	7	-4	82
87	1	C	14	24	33*	24	0	5	28		1	0,33	33	0,47	0,47	3	12	20	9	82
88	1	B	18	24*	29	20	0	9	48		1	0,24	24	0,43	0,43	3	12	21	9	82
89	1	D	22	24	34	15*	0	4	23		1	0,15	15	0,36	0,36	-7	3	12	0	82
90	1	B	11	31*	37	18	0	3	18		1	0,31	31	0,46	0,46	-13	-6	2	-9	82
91	1	C	14	28	38*	15	0	5	28		1	0,38	38	0,49	0,49	2	10	19	7	82
92	1	C	14	27	33*	21	0	5	28		1	0,33	33	0,47	0,47	3	12	21	9	82
93	1	B	17	30*	29	17	0	7	36		1	0,30	30	0,46	0,46	11	20	28	17	82
94	1	A	25*	30	27	13	0	4	19		1	0,25	25	0,43	0,43	0	10	19	7	82
95	1	B	11	33*	40	15	0	2	12		1	0,33	33	0,47	0,47	-3	6	15	3	82
96	1	A	15*	33	30	18	0	3	15		1	0,15	15	0,36	0,36	1	11	21	9	82
97	1	C	15	26	33*	23	0	4	20		1	0,33	33	0,47	0,47	14	23	31	19	82
98	1	C	16	26	33*	17	0	9	45		1	0,33	33	0,47	0,47	-7	2	10	-2	82
99	1	B	26	24*	29	16	0	5	28		1	0,24	24	0,43	0,43	-1	8	16	5	82
100	1	A	11*	37	33	15	0	5	24		1	0,11	11	0,31	0,31	3	13	23	11	82
101	1	D	12	28	39	16*	0	5	25		1	0,16	16	0,37	0,37	-6	3	11	0	82
102	1	C	15	31	30*	18	0	6	31		1	0,30	30	0,46	0,46	3	12	20	8	82
103	1	C	17	21	35*	20	0	7	38		1	0,35	35	0,48	0,48	1	9	17	6	82
104	1	C	21	17	34*	21	0	7	36		1	0,34	34	0,47	0,47	-1	7	16	4	82
105	1	C	11	32	36*	16	0	5	26		1	0,36	36	0,48	0,48	1	9	18	6	82
106	1	D	12	31	35	17*	0	5	27		1	0,17	17	0,37	0,37	1	10	19	7	82
107	1	A	16*	28	31	19	0	6	31		1	0,16	16	0,37	0,37	11	21	30	18	82
108	1	A	17*	22	32	21	0	8	39		1	0,17	17	0,38	0,38	-4	6	16	3	82
109	1	B	23	25*	31	16	0	5	25		1	0,25	25	0,43	0,43	9	18	27	15	82
110	1	B	9	37*	36	13	0	5	25		1	0,37	37	0,48	0,48	-3	5	13	2	82
111	1	D	14	32	31	17*	0	5	25		1	0,17	17	0,38	0,38	4	13	22	10	82
112	1	A	16*	27	34	18	0	5	28		1	0,16	16	0,36	0,36	10	21	31	18	82
113	1	C	18	29	27*	21	0	6	30		1	0,27	27	0,44	0,44	-4	5	14	2	82
114	1	B	21	30*	31	14	0	5	25		1	0,30	30	0,46	0,46	10	19	28	16	82
115	1	D	9	32	33	22*	0	4	23		1	0,22	22	0,41	0,41	13	23	32	20	82
116	1	B	13	32*	35	15	0	5	27		1	0,32	32	0,46	0,46	11	20	28	16	82
117	1	A	19*	25	34	16	0	7	34		1	0,19	19	0,39	0,39	22	32	41	29	82
118	1	C	16	29	30*	18	0	7	37		1	0,30	30	0,46	0,46	-8	0	8	-4	82
119	1	D	20	23	28	24*	0	6	30		1	0,24	24	0,43	0,43	14	23	32	21	82
120	1	B	9	35*	36	14	0	6	32		1	0,35	35	0,48	0,48	6	15	23	12	82
121	1	C	12	28	40*	15	0	6	30		1	0,40	40	0,49	0,49	6	14	22	11	82
122	1	C	11	27	36*	19	0	6	30		1	0,36	36	0,48	0,48	13	22	30	18	82
123	1	C	20	23	33*	18	0	7	35		1	0,33	33	0,47	0,47	3	12	21	9	82
124	1	D	20	23	27	24*	0	5	28		1	0,24	24	0,43	0,43	15	24	33	22	82
125	1	B	12	39*	33	12	0	5	26		1	0,39	39	0,49	0,49	10	18	26	15	82
126	1	C	15	29	36*	15	0	5	25		1	0,36	36	0,48	0,48	12	20	29	17	82
127	1	D	13	25	30	25*	0	7	35		1	0,25	25	0,43	0,43	15	24	32	21	82
128	1	B	18	28*	30	18	0	7	37		1	0,28	28	0,45	0,45	18	27	35	24	82
129	1	B	16	32*	31	14	0	7	36		1	0,32	32	0,47	0,47	18	27	35	24	82
130	1	D	14	30	32	18*	0	6	32		1	0,18	18	0,39	0,39	5	15	24	12	82
131	1	C	13	27	36*	17	0	7	36		1	0,36	36	0,48	0,48	10	19	27	15	82
132	1	D	13	25	31	24*	0	7	35		1	0,24	24	0,43	0,43	17	26	34	23	82
133	1	D	22	23	30	17*	0	8	42		1	0,17	17	0,38	0,38	2	12	22	9	82
134	1	D	24	24	26	19*	0	7	35		1	0,19	19	0,39	0,39	12	21	30	19	82
135	1	B	10	34*	34	16	0	6	33		1	0,34	34	0,47	0,47	5	13	22	10	82
136	1	D	16	33	30	15*	0	6	30		1	0,15	15	0,36	0,36	6	16	25	13	82
137	1	C	12	29	35*	18	0	6	31		1	0,35	35	0,48	0,48	-1	7	15	4	82
138	1	D	19	23	26	23*	0	8	44		1	0,23	23	0,42	0,42	15	25	34	22	82
139	1	C	20	22	33*	16	0	8	44		1	0,33	33	0,47	0,47	11	19	28	16	82
140	1	D	10	32	34	18*	0	7	35		1	0,18	18	0,38	0,38	9	18	27	15	82
141	1	A	16*	29	32	18	0	6	30		1	0,16	16	0,36	0,36	7	18	28	16	82
142	1	C	14	27	36*	17	0	6	33		1	0,36	36	0,48	0,48	2	11	19	7	82
143	1	D	21	26	28	17*	0	8	40		1	0,17	17	0,38	0,38	-5	2	9	-1	82
144	1	B	21	29*	26	18	0	7	36		1	0,29	29	0,45	0,45	8	17	26	14	82
145	1	B	11	30*	31	21	0	8	39		1	0,30	30	0,46	0,46	-6	2	10	-2	82
146	1	D	14	30	37	11*	0	8	42		1	0,11	11	0,32	0,32	-11	-4	2	-7	82
147	1	A	11*	30	29	23	0	8	41		1	0,11	11	0,31	0,31	-6	3	12	1	82
148	1	B	17	24*	31	18	0	10	52		1	0,24	24	0,43	0,43	-8	0	8	-3	82
149	1	D	22	27	25	17*	0	9	46		1	0,17	17	0,38	0,38	-8	-1	7	-3	82
150	1	A	10*	29	39	13	0	8	42		1	0,10	10	0,30	0,30	-5	4	14	2	

161	1	C	12	31	37*	12	0	8	39		1	0,37	37	0,48	0,48	-4	4	13	1	82				
162	1	A	16*	31	29	17	0	7	37		1	0,16	16	0,36	0,36	6	16	26	14	82				
163	1	C	20	22	32*	17	0	9	45		1	0,32	32	0,47	0,47	2	11	19	7	82				
164	1	C	21	23	32*	16	0	8	41		1	0,32	32	0,47	0,47	-5	3	12	0	82				
165	1	C	15	30	31*	17	0	8	40		1	0,31	31	0,46	0,46	-3	5	13	1	82				
166	1	B	12	32*	30	17	0	9	47		1	0,32	32	0,47	0,47	-2	7	16	4	82				
167	1	B	13	32*	30	17	0	8	43		1	0,32	32	0,47	0,47	4	12	20	9	82				
168	1	B	18	25*	29	15	0	13	65		1	0,25	25	0,43	0,43	3	12	21	9	82				
169	1	D	22	24	29	16*	0	9	48		1	0,16	16	0,36	0,36	4	13	23	11	82				
170	1	D	10	29	35	16*	0	10	51		1	0,16	16	0,37	0,37	5	14	24	12	82				
171	1	B	14	29*	33	15	0	10	50		1	0,29	29	0,45	0,45	3	12	20	8	82				
172	1	A	15*	30	30	15	0	10	51		1	0,15	15	0,36	0,36	8	19	29	16	82				
173	1	A	17*	24	31	16	0	12	63		1	0,17	17	0,38	0,38	-6	2	10	-1	82				
174	1	B	21	25*	28	17	0	10	50		1	0,25	25	0,43	0,43	1	10	19	7	82				
175	1	D	10	31	33	16*	0	9	49		1	0,16	16	0,37	0,37	0	9	18	7	82				
176	1	B	13	33*	32	13	0	9	45		1	0,33	33	0,47	0,47	8	16	25	13	82				
177	1	C	14	26	32*	19	0	8	43		1	0,32	32	0,47	0,47	-1	7	16	4	82				
178	1	B	18	24*	30	17	0	11	57		1	0,24	24	0,43	0,43	4	13	21	10	82				
179	1	C	24	24	27*	15	0	11	58		1	0,27	27	0,44	0,44	2	10	19	7	82				
180	1	C	10	29	37*	15	0	10	52		1	0,37	37	0,48	0,48	3	11	20	8	82				
181	1	D	13	32	33	13*	0	10	51		1	0,13	13	0,34	0,34	0	10	19	7	82				
182	1	A	14*	31	30	16	0	9	47		1	0,14	14	0,34	0,34	4	14	24	12	82				
183	1	B	21	21*	28	18	0	12	64		1	0,21	21	0,41	0,41	3	12	21	9	82				
184	1	A	32*	20	13	13	0	22	113		1	0,32	32	0,47	0,47	1	9	18	6	82				
185	1	C	11	26	29*	12	0	23	117		1	0,29	29	0,45	0,45	-5	4	12	1	82				
186	1	D	21	21	22	14*	0	22	115		1	0,14	14	0,35	0,35	4	14	24	12	82				
187	1	A	21*	20	18	17	0	24	125		1	0,21	21	0,40	0,40	4	13	21	10	82				

SubGroup number			: 0				SubTest number			: 0														
Number of persons in test			: 519				Number of selected items			: 187														
Minimum test score			: 0				Maximum test score			: 187														
Average test score			: 52,20				Standard deviation			: 13,97														
Average P-value			: 27,92				Std. Error of Measurement			: 5,95														
Average Rit			: 0,17																					
Coefficient Alpha			: 0,82				SE Coeff. Alpha			: 0,01														
GLB			: 0,00				Asymptotic GLB coef			: 0,00														
Items used in GLB proc			: 187																					
Cut-off score			: 60,5				Percentage failing			: 79,58														

Misclassifications:																								
Alpha based									GLB based															

-Rxx' case	Percentage		: 16,2				Percentage					: 40												
	Number		: 84				Number					: 208												
-Rxt case	Percentage		: 11,7				Percentage					: 40												
	Number		: 61				Number					: 208												

90% Confidence limits for Coefficient Alpha: (0,80 =< 0,82 =< 0,84)																								

Estimated Coefficient Alpha if this test had a standard norm length of 40 items: 0,49 (Spearman-Brown)																								

Cito, Measurement and Research Department. Arnhem, the Netherlands. © 2013.																								

Cito, Measurement and Research Department. Arnhem, the Netherlands. © 2013.

APPENDIX A3: TIAPLUS REPORT ON THE RESULTS OF THE ANALYSIS OF THE 60 ITEMS OF THE REF. NC-SES-TALA PART 1 AND 2

TiaPlus® Test and Item Analysis Build 314
Cito, Measurement and Research Department. Arnhem, the Netherlands. © 2013.

Population : Grade 12 Sesotho pupils
Test : TALA-based Sesotho2 (60-item) test
Date : Friday, 19 November 2021
Time : 17:01
Data file : C:\Albert Documents\ICELDA\ice02732_Sesotho2_60item_Tiaplus.txt
Database : C:\Cito\TiaPlus\TiaPlus2000.mdb
Missing handling : Missing as Zero

Persons: All persons

Items: All items

Test and Item Analysis

Item Label	Item nr.	Weight	Key	P- and A- values					% #					Weighted									
				A	B	C	D	E	F	O/D	Mis-singl	Max	Mean	P	Sd	R\$K	[Lo	Rit	Hi]	Rir	AR		
	1	1	D	20	29	18	18*	13		2	10	1	0,18	18	0,39	0,39	0	9	17	3	77		
	2	1	E	16	26	28	17	10*		3	17	1	0,10	10	0,30	0,30	18	27	36	23	76		
	3	1	A	24*	24	24	17	7		3	17	1	0,24	24	0,43	0,43	-2	7	16	1	77		
	4	1	C	24	20	24*	20	8		3	16	1	0,24	24	0,43	0,43	3	11	20	6	77		
	5	1	B	21	22*	24	16	12		5	27	1	0,22	22	0,42	0,42	-9	-1	7	-7	77		
	6	1	D	14	20	25	38*	2		1	7	1	0,38	38	0,49	0,49	28	36	43	30	76		
	7	1	C	16	18	44*	17	3		3	14	1	0,44	44	0,50	0,50	37	43	49	37	76		
	8	1	C	20	16	45*	17	1		1	4	1	0,45	45	0,50	0,50	35	42	48	36	76		
	9	1	D	21	21	22	34*	0		2	10	1	0,34	34	0,47	0,47	35	43	50	37	76		
	10	1	D	19	21	20	36*	2		2	11	1	0,36	36	0,48	0,48	33	41	48	35	76		
	11	1	C	20	21	38*	14	3		3	17	1	0,38	38	0,49	0,49	38	45	51	39	76		
	12	1	B	15	43*	27	12	1		2	12	1	0,43	43	0,50	0,50	26	34	42	28	76		
	13	1	C	18	21	39*	16	2		3	18	1	0,39	39	0,49	0,49	39	45	52	40	76		
	14	1	C	27	22	35*	14	1		1	7	1	0,35	35	0,48	0,48	49	55	60	50	75		
	15	1	C	16	23	45*	13	2		2	8	1	0,45	45	0,50	0,50	38	44	51	38	76		
	16	1	C	20	23	34*	18	2		2	12	1	0,34	34	0,48	0,48	30	38	45	32	76		
	17	1	D	28	22	20	26*	2		2	8	1	0,26	26	0,44	0,44	30	39	47	33	76		
	18	1	B	13	44*	27	13	0		2	11	1	0,44	44	0,50	0,50	26	34	41	27	76		
	19	1	B	26	38*	18	13	3		3	14	1	0,38	38	0,49	0,49	31	39	46	33	76		
	20	1	C	30	19	35*	13	1		2	8	1	0,35	35	0,48	0,48	31	39	46	33	76		
	21	1	C	18	24	36*	16	2		4	22	1	0,36	36	0,48	0,48	31	38	45	32	76		
	22	1	C	27	24	33*	12	1		3	14	1	0,33	33	0,47	0,47	29	37	44	31	76		
	23	1	D	10	31	28	27*	1		3	15	1	0,27	27	0,44	0,44	29	38	46	32	76		
	24	1	B	22	35*	20	17	2		4	20	1	0,35	35	0,48	0,48	25	33	41	27	76		
	25	1	A	31*	25	29	10	1		4	21	1	0,31	31	0,46	0,46	23	32	40	26	76		
	26	1	B	27	24*	32	14	0		2	11	1	0,24	24	0,43	0,43	8	17	26	11	77		
	27	1	C	14	24	33*	24	0		5	28	1	0,33	33	0,47	0,47	8	17	25	10	77		
	28	1	C	14	28	38*	15	0		5	28	1	0,38	38	0,49	0,49	9	17	25	10	77		
	29	1	C	14	27	33*	21	0		5	28	1	0,33	33	0,47	0,47	4	13	22	7	77		
	30	1	B	17	30*	29	17	0		7	36	1	0,30	30	0,46	0,46	15	23	32	17	76		
	31	1	C	15	26	33*	23	0		4	20	1	0,33	33	0,47	0,47	18	27	35	20	76		
	32	1	C	15	31	30*	18	0		6	31	1	0,30	30	0,46	0,46	4	13	21	7	77		
	33	1	C	18	29	27*	21	0		6	30	1	0,27	27	0,44	0,44	3	11	19	5	77		
	34	1	B	21	30*	31	14	0		5	25	1	0,30	30	0,46	0,46	11	20	29	14	77		
	35	1	D	9	32	33	22*	0		4	23	1	0,22	22	0,41	0,41	18	28	37	22	76		
	36	1	A	19*	25	34	16	0		7	34	1	0,19	19	0,39	0,39	22	32	41	27	76		
	37	1	D	20	23	28	24*	0		6	30	1	0,24	24	0,43	0,43	21	30	38	24	76		
	38	1	C	12	28	40*	15	0		6	30	1	0,40	40	0,49	0,49	12	20	28	13	77		
	39	1	C	11	27	36*	19	0		6	30	1	0,36	36	0,48	0,48	19	27	35	21	76		
	40	1	D	20	23	27	24*	0		5	28	1	0,24	24	0,43	0,43	22	31	39	25	76		
	41	1	B	12	39*	33	12	0		5	26	1	0,39	39	0,49	0,49	16	24	32	17	76		
	42	1	D	13	25	30	25*	0		7	35	1	0,25	25	0,43	0,43	17	26	35	20	76		
	43	1	B	18	28*	30	18	0		7	37	1	0,28	28	0,45	0,45	21	29	37	24	76		
	44	1	B	16	32*	31	14	0		7	36	1	0,32	32	0,47	0,47	21	30	37	23	76		
	45	1	C	13	27	36*	17	0		7	36	1	0,36	36	0,48	0,48	16	24	32	17	76		
	46	1	D	13	25	31	24*	0		7	35	1	0,24	24	0,43	0,43	17	26	36	21	76		
	47	1	D	24	24	26	19*	0		7	35	1	0,19	19	0,39	0,39	14	24	32	18	76		
	48	1	D	19	23	26	23*	0		8	44	1	0,23	23	0,42	0,42	20	29	38	24	76		
	49	1	C	21	23	30*	16	0		10	50	1	0,30	30	0,46	0,46	6	15	23	8	77		
	50	1	C	11	31	32*	16	0		9	45	1	0,32	32	0,47	0,47	5	13	22	7	77		
	51	1	C	18	22	33*	17	0		10	50	1	0,33	33	0,47	0,47	4	13	21	6	77		
	52	1	C	29	23	30*	12	0		7	35	1	0,30	30	0,46	0,46	3	11	19	5	77		
	53	1	A	16*	31	29	17	0		7	37	1	0,16	16	0,36	0,36	7	17	27	12	77		
	54	1	C	20	22	32*	17	0		9	45	1	0,32	32	0,47	0,47	9	17	25	11	77		
	55	1	C	21	23	32*	16	0		8	41	1	0,32	32	0,47	0,47	0	8	16	2	77		
	56	1	C	15	30	31*	17	0		8	40	1	0,31	31	0,46	0,46	0	8	16	1	77		
	57	1	B	13	33*	32	13	0		9	45	1	0,33	33	0,47	0,47	8	16	25	10	77		
	58	1	C	14	26	32*	19	0		8	43	1	0,32	32	0,47	0,47	4	13	21	6	77		
	59	1	B	18	24*	30	17	0		11	57	1	0,24	24	0,43	0,43	7	16	24	10	77		
	60	1	C	24	24	27*	15	0		11	58	1	0,27	27	0,44	0,44	6	14	22	8	77		

SubGroup number : 0
Number of persons in test : 519
Minimum test score : 0
Average test score : 18,57
SubTest number : 0
Number of selected items : 60
Maximum test score : 60
Standard deviation : 7,12

Average P-value	: 30,95	Std. Error of Measurement	: 3,44
Average Rit	: 0,26		
Coefficient Alpha	: 0,77	SE Coeff. Alpha	: 0,01
GLB	: 0,00	Asymptotic GLB coef	: 0,00
Items used in GLB proc	: 60		
Cut-off score	: 19,5	Percentage failing	: 62,43

Misclassifications:

	Alpha based		GLB based
-Rxx' case	Percentage : 22	Percentage	: 49,5
	Number : 114	Number	: 257
-Rxt case	Percentage : 15,9	Percentage	: 49,5
	Number : 83	Number	: 257

 90% Confidence limits for Coefficient Alpha: (0,74 =< 0,77 =< 0,79)

Estimated Coefficient Alpha if this test had a standard
 norm length of 40 items: 0,69 (Spearman-Brown)

Cito, Measurement and Research Department. Arnhem, the Netherlands. © 2013.

APPENDIX B1: ETHICS CERTIFICATE



Private Bag X1290,
Potchefstroom South Africa
2520

Tel: 018 299-1111/2222
Fax: 018 299-4910
Web: <http://www.nwu.ac.za>

Research Ethics Regulatory Committee

Tel: 018 299-4849
Email: nkosinathi.machine@nwu.ac.za

28 June 2019

ETHICS APPROVAL LETTER OF STUDY

Based on approval by the **Language Matters Ethics Committee (LM-REC)** on 28/06/2019, the Language Matters Research Ethics Committee hereby **approves** your study as indicated below. This implies that the North-West University Research Ethics Regulatory Committee (NWU-RERC) grants its permission that, provided the special conditions specified below are met and pending any other authorisation that may be necessary, the study may be initiated, using the ethics number below.

Study title: Equalizing Grade 12 home language examination papers in a multilingual education department: The case of Sesotho and English.

Study Leader/Supervisor (Principal Investigator)/Researcher: Prof. G. Butler

Student/Researcher: J.T Mahlasela

Ethics number:

N	W	U	-	0	0	6	2	0	-	1	9	-	A	7
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Institution Study Number Year Status

Status: S = Submission; R = Re-Submission; P = Provisional Authorisation; A = Authorisation

Application Type: Single Study

Commencement date: 28/06/2019

Expiry date: 27/06/2020

Risk Category: Low

Approval of the study is initially provided for a year, after which continuation of the study is dependent on receipt and review of the annual (or as otherwise stipulated) monitoring report and the concomitant issuing of a letter of continuation.

Special in process conditions of the research for approval (if applicable):

- Translation of the informed consent document to the languages applicable to the study participants should be submitted to the LM-REC (if applicable).
- Any research at governmental or private institutions, permission must still be obtained from relevant authorities and provided to the LM-REC. Ethics approval is required BEFORE approval can be obtained from these authorities.

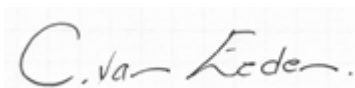
General conditions:

While this ethics approval is subject to all declarations, undertakings and agreements incorporated and signed in the application form, the following general terms and conditions will apply:

- *The study leader/supervisor (principle investigator)/researcher must report in the prescribed format to the LM-REC:*
 - *annually (or as otherwise requested) on the monitoring of the study, whereby a letter of continuation will be provided, and upon completion of the study; and*
 - *without any delay in case of any adverse event or incident (or any matter that interrupts sound ethical principles) during the course of the study.*
- *The approval applies strictly to the proposal as stipulated in the application form. Should any amendments to the proposal be deemed necessary during the course of the study, the study leader/researcher must apply for approval of these amendments at the LM-REC, prior to implementation. Should there be any deviations from the study proposal without the necessary*
 - *request access to any information or data at any time during the course or after completion of the study;*
 - *to ask further questions, seek additional information, require further modification or monitor the conduct of your research or the informed consent process;*
 - *withdraw or postpone approval if:*
 - *any unethical principles or practices of the study are revealed or suspected;*
 - *it becomes apparent that any relevant information was withheld from the LM-REC or that information has been false or misrepresented;*
 - *submission of the annual (or otherwise stipulated) monitoring report, the required amendments, or reporting of adverse events or incidents was not done in a timely manner and accurately; and / or*
 - *new institutional rules, national legislation or international conventions deem it necessary.*
- *LM-REC can be contacted for further information or any report templates via Bertus.VanRooy@nwu.ac.za or 018 210 3482.*

The LM-REC would like to remain at your service as scientist and researcher, and wishes you well with your study. Please do not hesitate to contact the LM-REC or the NWU-RERC for any further enquiries or requests for assistance.

Yours sincerely



Prof C Van Eeden
Chairperson NWU Language Matters Research Ethics Committee

Original details: (22351930) C:\Users\22351930\Desktop\ETHICS APPROVAL LETTER
OF STUDY.docm 8 November 2018

File reference: 9.1.5.4.2

APPENDIX B2: DBE APPROVALS

Enquiries: KK Motshumi
Ref: Research Permission: JT Mahlasela assisted by Phaladi Fakude, Lebohang Mathibela and Lebona Mafia
Tel. 051 404 9283 / 9221 / 079 503 4943 Email: K.Motshumi@fseducation.gov.za



Dear Mr. Mahlasela assisted by Pheladi Fakude, Lebohang Mathibela

APPROVAL TO CONDUCT RESEARCH IN THE FREE STATE DEPARTMENT OF EDUCATION

1. This letter serves as an acknowledgement of receipt of your request to conduct research in the Free State Department of Education.

Research Topic: Equalizing grade 12 home language examination papers in a multilingual education department: The case of English and Sesotho.

Schools: Iketsetseng, Cedar, Kahobotjha and Nkgopoleng secondary schools.

2. **Target Population:** 708 Sesotho grade 12 home language learners to complete standardized tests.

3. **Period of research:** From date of signature until 30 September 2019. Please note the department does not allow any research to be conducted during the fourth term (quarter) of the academic year. Should you fall behind your schedule by three months to complete your research project in the approved period, you will need to apply for an extension.

4. The approval is subject to the following conditions:

- 4.1 The collection of data should not interfere with the normal tuition time or teaching process.
- 4.2 A bound copy of the research document or a CD, should be submitted to the Free State Department of Education, Room 319, 3rd Floor, Old CNA Building, Charlotte Maxeke Street, Bloemfontein.
- 4.3 You will be expected, on completion of your research study to make a presentation to the relevant stakeholders in the Department.
- 4.4 The ethics documents must be adhered to in the discourse of your study in our department.
- 4.5 Please note that costs relating to all the conditions mentioned above are your own responsibility.

Your Faithful
DR J. NE
CHIEF FINANCIAL OFFICER

DATE: 25/6/2019

RESEARCH APPLICATION JT MAHLASELA ASSISTED BY PHALADI FAKUDE, LEBOHANG MATHIBELA AND LEBONA MAFIA
PERMISSION 07 05 2019 STRATEGIC PLANNING, POLICY & RESEARCH DIRECTORATE
Private Bag X20565, Bloemfontein, 9300 - Room 318, Old CNA Building, 3rd Floor, Charlotte Maxeke Street, Bloemfontein
Tel: (051) 404 9283 / 9221 Fax: (086) 6678 678

Ref: Notification of research: JT Mahlasela Assisted By Phaladi Fakude,
Lebohang Mathibela and Lebona Mafia
Tel. 051 404 9221 / 079 503 4943
Email: K. Motshumi@fseclucation.gov.za



education
Department of
Education
FREE STATE PROVINCE

The Director: Fezile Dabi District

Dear Ms. Chuta

**NOTIFICATION TO CONDUCT RESEARCH PROJECT IN YOUR DISTRICT BY JT MAHLASELA
ASSISTED BY PHALADI FAKUDE, LEBOHANG MATHIBELA AND LEBONA MAFIA**

1. The above mentioned candidate was granted permission to conduct research in your district as follows:

Research Topic: Equalizing grade 12 home language examination papers in a multilingual education department: The case of English and Sesotho.

Schools: Iketseseng, Cedar, Kahobotiha and Nkqopoleng secondary schools.

2. **Target Population:** 708 Sesotho grade 12 home language learners to complete standardized tests.
3. **Period:** From date of signature until the 30th September 2019. Please note the department does not allow any research to be conducted during the fourth / academic quarter of the year nor during normal school hours.
4. **Research benefits:** The study intends to address the issue of equity and Justice across all grade 12 Home language papers including those of the provincial papers such as the Free State.
5. Logistical procedures were met, in particular ethical considerations for conducting research in the Free State Department of Education.
6. Strategic Planning, Policy and Research Directorate will make the necessary arrangements for the researchers to present the findings and recommendations to the relevant officials in the district.

Yours sincerely


R JEM SEKOLANYANE
HIEF FINANCIAL OFFICER

ATE: 25/6/2019

RESEARCH APPLICATION FOR JT MAHLASELA ASSISTED BY PHALADI FAKUDE, LEBOHANG MATHIBELA AND LEBONA MAFIA NOTIFICATION

07 05 2019. FEZILE DABI DISTRICT

Strategic Planning, Research & Policy Directorate

Private Bag X20565, Bloemfontein, 9300 - Old CNA Building, Room 318, 3rd Floor, Charlotte Mexeke Street, Bloemfontein

Tel: (051) 404 9283 / 9221 Fax: (086) 6678 678

APPENDIX B3: CONSENT AND ASSENT FORMS



P.O.BOX 1174

VANDERBIJLPARK

1900

DATE: 07 March 2019

Tel: (016) 910 3215

TO WHOM IT MAY CONCERN (REQUEST FOR CONSENT)

Dear Sir/Madam

Re: Parent/Guardian – Request for permission to conduct research

Request is hereby made to conduct research on your child as part of the school population. This research is based on the notion that Grade 12 Home language examination papers are not equal and need to be equalised. Your child will be part of the sample upon which a standardised test will be conducted in order to test two strategies of equalising Grade 12 exam papers. This test will be based purely on school assessment procedure and no physical tests will be conducted upon children.

Your child's participation in this research is voluntary and may choose to withdraw from this process any time. If a child chooses to participate or withdraw from the research process, there will be no penalty for doing so. Written consent will be obtained from your child.

For the purposes of anonymity and confidentiality, the names of your child will not be mentioned throughout the data findings and results of this research.

Should you have any questions or enquiries about this research, please do not hesitate to contact me or my supervisor at the above mentioned address and phone number.

Thanking you in advance

Yours faithfully

Johannes Tsietsi Mahlasela

Signature _____ Date: _____

P.O.BOX 1174

VANDERBIJLPARK

1900

DATE: 07 March 2019

Tel: (016) 910 3215

LEARNER ASSENT FORM

Learner Number: _____

Name of School: _____

Date: _____

Information and assent:

TALA is a Test of Advanced Language Ability meant to test the standard of Home Language Examination papers through the eleven official languages of South Africa.

The results of this test will help the researcher to make informed decision about the standard of Sesotho Home Language papers and make recommendations to the authorities.

By signing this form you give the researcher permission to use the results of this test for further research and recommendations without using your name.

I, (Please write your full names and surname) _____

Give permission to the researcher, J.T.Mahlasela and members of his research team to report the data gathered in this test in my PhD, books, academic journals, academic conferences and institutional reports.

Signature: _____ **Date:** _____