

# **Die teoretiese begroning vir die ontwerp van 'n nagraadse toets van akademiese geletterdheid in Afrikaans**

Gini Keyser

Hierdie verhandeling word voorgelê ingevolge die vereistes van die graad MA (met spesialisasie in Afrikaans en Nederlands) in die Departement Afrikaans en Nederlands, Duits en Frans in die Fakulteit Geesteswetenskappe aan die Universiteit van die Vrystaat

Julie 2017

Studieleier: Prof. A.J. Weideman

## **Verklaring**

Ek verklaar dat die verhandeling wat hiermee ingedien word vir die graad MA (met spesialisasie in Afrikaans en Nederlands) aan die Universiteit van die Vrystaat my eie, onafhanklike werk is en nie voorheen deur my vir 'n graad aan 'n universiteit of fakulteit ingedien is nie. Ek doen voorts afstand van die outeursreg in die verhandeling ten gunste van die Universiteit van die Vrystaat.

Naam: Gini Keyser

Datum: Julie 2017

## Dankwoord

Allereers bedank ek my Hemelse Vader. Dankie, Here, vir genade onbeskryflik groot. Met hierdie studie bely ek my afhanklikheid van u voorsienigheid.

Hiermee bedank ek ook die volgende persone en instansies sonder wie se hulp en ondersteuning die uitvoering van hierdie studie nie moontlik sou wees nie:

- my studieleier en mentor, prof. A.J. Weideman, vir sy kundigheid, geduld en volgehoue motivering. So ook sy vrou, Anna Weideman, vir haar vriendelikheid en gasvryheid wanneer ek hulle besoek het.
- my ouers wat my die geleentheid gegee het om verder te kan studeer en vir hulle vertrou in my vermoëns.
- die Suid-Afrikaanse Akademie vir Wetenskap en Kuns (SAAWK) vir hulle geldelike ondersteuning in die vorm van 'n beurs.
- die Inter-Institutional Centre for Language Development and Assessment (ICELDA) vir die beskikbaarstelling van fondse ten einde die navorsing te kon uitvoer.
- verskeie kollegas verbonde aan die Universiteit van die Vrystaat, Universiteit Stellenbosch, Noordwes-Universiteit, Universiteit van Pretoria en Aros vir hulle bystand met die loodsing van die toets wat as deel van hierdie studie ontwerp is.
- my kollegas by die Eenheid vir Taalontwikkeling aan die Universiteit van die Vrystaat vir hulle belangstelling en begrip, in die besonder dr. Annette de Wet.
- my goeie vriende vir hulle gebede en aanmoediging onder alle omstandighede.
- my twee Boston Terriërs, Kallie en Dirkie, wat my geselskap gehou het in my studeerkamer.

“En gelukkig is elkeen wat nie aan My begin twyfel nie.”

Mattheus 10:6

**Vir my pa, Dawie en my ma, Sunita**

## Inhoudsopgawe

Hoofstuk 1: Inleiding .....	1
1. Rasionaal vir die studie .....	1
1.1 Die problematiek ten opsigte van die akademiese geletterdheid van voorgaadse studente .....	1
1.2 Gebrekkige akademiese geletterdheid: Die nagraadse nadraai .....	5
1.3 Die behoefte aan 'n nagraadse toets van akademiese geletterdheid in Afrikaans .....	8
2. Doel van die studie .....	12
3. Navorsingsmetodologie .....	17
4. Waarde van die studie .....	20
5. Uiteensetting van die studie .....	22
Hoofstuk 2: Verantwoordelike toetsontwerp: 'n Teoretiese vertrekpunt .....	24
1. Inleiding .....	24
2. Die toegepaste linguistiek as 'n dissipline van ontwerp .....	25
2.1 'n Definisie van die toegepaste linguistiek: Moderne en postmoderne opvattinge .....	25
2.2 'n Oorsig van die verskillende paradigmas binne die toegepaste linguistiek .....	30
2.3 Ontwerp as die wesenskenmerk van die toegepaste linguistiek ....	34
3. Taaltoetsing: 'n Subdissipline van die toegepaste linguistiek .....	38
3.1 Die ontwikkelingsgang van taaltoetsing .....	38
3.2 Geldigheid in taaltoetsing: Die tradisionele versus hedendaagse perspektief .....	45
3.3 'n Alternatiewe perspektief: 'n Balans tussen die konstitutiewe en regulatiewe voorwaardes vir toetsontwerp .....	54
4. Slot .....	63

Hoofstuk 3: Die ontwerp van 'n nagraadse toets van akademiese geletterdheid in Afrikaans.....	64
1. Inleiding .....	64
2. Die keuse van 'n toetskonstruk: 'n Hersiene konstruk van akademiese geletterdheid as teoretiese rasionaal.....	66
2.1 Die toetskonstruk van TALL, TAG en TALPS .....	66
2.2 'n Verfynde konstruk as basis vir die ontwerp van TAGNaS.....	76
3. TAGNaS as 'n potensieële driefasetoets .....	82
4. Toetsspesifikasies van TAGNaS.....	89
5. Die subtoetse van TAGNaS .....	92
6. Die gebruik van veelkeuse-items in TAGNaS.....	103
7. Die keuse van 'n tema en leestekste vir TAGNaS .....	104
8. Die woordeskatitems in TAGNaS en die <i>Akademiese Woordelys vir Afrikaans</i> .....	108
9. Die skryf van die items in TAGNaS .....	110
10. Slot.....	111
Hoofstuk 4: Die loodsing en verfyning van 'n nagraadse toets van akademiese geletterdheid in Afrikaans.....	113
1. Inleiding .....	113
2. Die toetsontwikkelingproses voortgesit: Die eerste loodstoetsing van TAGNaS.....	114
2.1 Die nodigheid en belang van loodstoetsing .....	114
2.2 Getalle en deelnemers .....	115
2.3 Prosedure en benadering .....	118
2.4 Produktiwiteit en parameters.....	122
3. Die eerste loodstoetsing van TAGNaS: Analise en bespreking van toetsresultate per subtoets .....	133
3.1 Skommeltekste .....	133
3.2 Interpretasie van visuele inligting .....	137
3.3 Akademiese woordeskat.....	141
3.4 Tekstipe .....	143

3.5	Teksbegrip en Vergelyking tussen tekste.....	145
3.6	Grammatika en teksverband.....	149
3.7	Opsommingtaak .....	151
3.8	Verwysings.....	153
4.	Die verfyningproses: 'n Seleksie van produktiewe toetsitems.....	154
5.	Slot .....	158
Hoofstuk 5: Die verdere verfyning van 'n nagraadse toets van akademiese geletterdheid in Afrikaans: Resultate van 'n tweede loodstoets..		
1.	Inleiding .....	160
2.	'n Volgende stap in die toetsontwikkelingsproses: Die tweede loodstoetsing van TAGNaS .....	161
2.1	Getalle en deelnemers .....	161
2.2	Prosedure, benadering en parameters.....	162
3.	Die tweede loodstoetsing van TAGNaS: Analise en bespreking van toetsresultate.....	163
3.1	Statistiese resultate van eerste toetsfase (verpligte/opsionele siftingstoets) .....	163
3.2	Statistiese resultate van tweede toetsfase .....	170
4.	Slot .....	179
Hoofstuk 6: Gevolgtrekking .....		
1.	Doel van die studie: 'n Samevattende terugblik .....	181
1.1	Die ontwerp, loodsing en verfyning van TAGNaS.....	182
1.2	Die teoretiese begroning vir die ontwerp van TAGNaS .....	183
1.3	Die eksperimentering met nuwe en/of gewysigde subtoetse/taaktipes.....	191
1.4	Die moontlike implementering van 'n driefasige toetsformaat ...	191
2.	Beperkings van die studie .....	192
3.	Aanbevelings vir verdere navorsing .....	194
Bronnelys .....		197

Abstrak .....	212
Bylaag A: Toetsspesifikasies vir die Test of Academic Literacy for Postgraduate Students (TALPS) .....	216
Bylaag B: Toetsspesifikasies vir die Toets van Akademiese Geletterdheid vir Nagraadse Studente (TAGNaS) .....	218
Bylaag C: Merkrubriek (TALPS) .....	223
Bylaag D: Vertaalde merkrubriek (TAGNaS) .....	224
Bylaag E: Potensiële probleemitems in die verfynde weergawe van TAGNaS .....	225



## Lys van tabelle

Tabel 2.1	Sewe opeenvolgende tradisies binne die toegepaste linguistiek	31
Tabel 2.2	Messick se fasette van geldigheid	49
Tabel 2.3	McNamara en Roever (2006) se herinterpretasie van Messick se fasette van geldigheid	51
Tabel 2.4	Bachman en Palmer (1996) se model van toetsbruikbaarheid	51
Tabel 2.5	Die verband tussen enkele fundamentele oorwegings in taaltoetsing	54
Tabel 2.6	Vereistes vir taaltoetsing	54
Tabel 2.7	Konstitutiewe en regulatiewe momente in toegepaste linguistiese ontwerpe	61
Tabel 2.8	Beginsels vir verantwoordelike toetsontwerp	62
Tabel 3.1	Twee sieninge van taal	68
Tabel 3.2	Bachman en Palmer se konstruk	71
Tabel 3.3	TAGNaS as 'n potensiële driefasetoets	83
Tabel 3.4	TAGNaS as 'n potensiële driefasetoets (met subtoetse)	88
Tabel 3.5	Betroubaarheidswaardes van die subtoetse in TALPS soos afgelê in 2011 en 2012 by die Universiteit van die Vrystaat (UFS)	91
Tabel 3.6	Beskrywende statistiek vir die subtoetse in TALPS soos afgelê in 2011 en 2012 by die Universiteit van die Vrystaat (UFS)	91
Tabel 3.7	'n Voorbeeld van visuele analise (Patterson & Weideman, 2013b:143)	99
Tabel 3.8	Matriks vir beoordelaars se kommentaar oor 'n teks	106
Tabel 3.9	Skedule vir die skryf van items vir TAGNaS	111
Tabel 4.1	Eerste loodstoetsing van TAGNaS: 'n Oorsig van getalle (loodstoetsitems, -kandidate en geselekteerde items)	117

Tabel 4.2	'n Matriks vir die beoordeling van toetsitems	122
Tabel 4.3	Beskrywende statistiek vir skommeltekste soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)	134
Tabel 4.4	Betroubaarheidskoeffisiënte van skommeltekste soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)	135
Tabel 4.5	Interne korrelasies tussen die items van Skommelteks 1 soos bepaal deur TiaPlus (Eerste loodstoetsing)	135
Tabel 4.6	Interne korrelasies tussen die items van Skommelteks 2 soos bepaal deur TiaPlus (Eerste loodstoetsing)	135
Tabel 4.7	“Item fit”-statistiek vir die items van Skommelteks 1 soos bepaal deur Winsteps (Eerste loodstoetsing)	137
Tabel 4.8	Geïdentifiseerde probleemitems: Interpretasie van visuele inligting (Eerste loodstoetsing)	137
Tabel 4.9	Itemstatistiek van probleemitems: Interpretasie van visuele inligting (Eerste loodstoetsing)	138
Tabel 4.10	Konvergensietabel (Deel 1 en 2): Interpretasie van visuele inligting (Eerste loodstoetsing)	139
Tabel 4.11	Beskrywende statistiek vir die akademiese woordeskatsubtoets soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)	142
Tabel 4.12	Betroubaarheidskoeffisiënte van die akademiese woordeskat-taaktipes soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)	143
Tabel 4.13	Itemstatistiek van Tekstipe: Weergawe 1 (Eerste loodstoetsing)	144
Tabel 4.14	Itemstatistiek van Tekstipe: Weergawe 2 (Eerste loodstoetsing)	144
Tabel 4.15	Beskrywende statistiek vir die Teksbegrip-subtoets soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)	146

Tabel 4.16	Betroubaarheidskoeffisiënte van die Teksbegrip-loodstoetse soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)	146
Tabel 4.17	Beskrywende statistiek vir die Teksbegrip-subtoets soos bepaal deur TiaPlus (Eerste loodstoetsing)	146
Tabel 4.18	Itemstatistiek van die vergelykende items in die Teksbegrip-loodstoetse (Eerste loodstoetsing)	149
Tabel 4.19	Beskrywende statistiek vir die Grammatika en teksverband-subtoets soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)	150
Tabel 4.20	Betroubaarheidskoeffisiënte van Grammatika en teksverband 1a en b soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)	150
Tabel 4.21	Betroubaarheidskoeffisiënte van Grammatika en teksverband 2a en b soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)	150
Tabel 4.22	Beskrywende statistiek van Opsomming: Weergawe 1 soos bepaal deur TiaPlus (Eerste loodstoetsing)	151
Tabel 4.23	Beskrywende statistiek van Opsomming: Weergawe 2 soos bepaal deur TiaPlus (Eerste loodstoetsing)	152
Tabel 4.24	“Item fit”-statistiek vir die items van Opsomming, Weergawe 2 soos bepaal deur Winsteps (Eerste loodstoetsing)	152
Tabel 4.25	Beskrywende statistiek van die Verwysing-subtoets soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)	153
Tabel 4.26	Betroubaarheidskoeffisiënte van die Verwysing-subtoets soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)	153
Tabel 4.27	Itemstatistiek van die geselekteerde items vir die verfynde toetsweergawe van TAGNaS	154- 158
Tabel 5.1	Beskrywende statistiek vir Toetsfase 1 soos bepaal deur Iteman 4.3	165

Tabel 5.2	Betroubaarheidskoeffisiënte van die subtoetse in Toetsfase 1 soos bepaal deur Iteman 4.3	166
Tabel 5.3	Interne korrelasies tussen die subtoetse van Toetsfase 1 soos bepaal deur Iteman 4.3	167
Tabel 5.4	Interne korrelasies tussen die subtoetse van Toetsfase 1 en die toets as geheel soos bepaal deur TiaPlus	167
Tabel 5.5	Potensiële misklassifikasies van Toetsfase 1 soos bepaal deur TiaPlus	169
Tabel 5.6	Beskrywende statistiek vir Toetsfase 2 soos bepaal deur Iteman 4.3	172
Tabel 5.7	Betroubaarheidskoeffisiënte van die subtoetse in Toetsfase 2 soos bepaal deur Iteman 4.3	174
Tabel 5.8	Interne korrelasies tussen die subtoetse van Toetsfase 2 soos bepaal deur Iteman 4.3	175
Tabel 5.9	Interne korrelasies tussen die subtoetse van Toetsfase 2 en die toets as geheel soos bepaal deur TiaPlus	176
Tabel 5.10	Potensiële misklassifikasies van Toetsfase 2 soos bepaal deur TiaPlus	178

## Lys van figure

Figuur 2.1	Terminale funksies van 'n toegepaste linguistiese ontwerp	36
Figuur 2.2	Vyf fases van toegepaste linguistiese ontwerpe	37
Figuur 2.3	Konstitutiewe konsepte en regulatiewe idees in toegepaste linguistiese ontwerpe	58
Figuur 3.1	Terminale funksies van akademiese diskoers	78
Figuur 4.1	Faktoranalise: Interpretasie van visuele inligting (Eerste loodstoetsing)	139
Figuur 4.2	Persoon-item-passing: Interpretasie van visuele inligting (Eerste loodstoetsing)	141
Figuur 4.3	Faktoranalise: Teksbegrip 1 (Eerste loodstoetsing)	147
Figuur 4.4	Faktoranalise: Teksbegrip 2 (Eerste loodstoetsing)	148
Figuur 4.5	Faktoranalise: Teksbegrip 3 (Eerste loodstoetsing)	148
Figuur 5.1	Frekwensiedistribusie van totale toetstellings (Toetsfase 1) soos bepaal deur Iteman 4.3	163
Figuur 5.2	Faktoranalise: Woordeskat – Enkelgaping (Tweede loodstoetsing)	168
Figuur 5.3	Faktoranalise: Grammatika en teksverband (Tweede loodstoetsing)	169
Figuur 5.4	Frekwensiedistribusie van totale toetstellings (Toetsfase 2) soos bepaal deur Iteman 4.3	171
Figuur 5.5	Faktoranalise: Woordeskat Enkelgaping (Tweede loodstoetsing)	177
Figuur 5.6	Faktoranalise: Teksbegrip (Tweede loodstoetsing)	177
Figuur 5.7	Faktoranalise: Vergelyking tussen tekste (Tweede loodstoetsing)	178

# Hoofstuk 1

## Inleiding

### 1. Rasionaal vir die studie

#### 1.1 Die problematiek ten opsigte van die akademiese geletterdheid van voorgraadse studente

Lae deurvloeikoerse en hoë uitvalsyfers bly 'n kommerwekkende tendens in die Suid-Afrikaanse hoërondewyssektor. Eerstejaarstudente wat ontoereikend vir die uitdagings van die tersiêre leeromgewing voorberei is, en heel waarskynlik nie met welslae hulle kwalifikasies binne die minimum voorgeskrewe studiejare sal verwerf nie, is 'n omstandigheid waarvan gereeld in die literatuur melding gemaak word (Van Dyk, 2010:1; Van Dyk & Van der Slik, 2012:45; Van Rooy & Coetzee-Van Rooy, 2015). In 2007, byvoorbeeld, verskyn 'n verslag van die Raad vir Hoër Onderwys (RHO) waarin aangetoon word dat, teen die einde van 2004, slegs 30% van alle voorgraadse studente aan tersiêre onderwysinrigtings in Suid-Afrika daarin kon slaag om binne 'n tydperk van vyf jaar te gradueer. Boonop het 56% van dié studente die opvoedkundige instelling waarby hulle aanvanklik as eerstejaar ingeskryf het, verlaat sonder om 'n graad te behaal, terwyl 14% ten tyde van die ondersoek steeds as voorgraadse studente geregistreer was. Die geskatte deurvloeikoerse van hierdie groep studente word as 44% aangegee (Scott, Yeld & Hendry, 2007). Uit 'n meer resente publikasie van die RHO (Access and throughput in South African higher education: Three case studies, 2010), waarin die kwessie van studentetoeegang en -sukses opnuut in oënskou geneem word, is dit duidelik dat die uitdaging van onbevredigende studenteprestasie 'n durende een is. In hierdie studie wat die deurvloeikoerse by drie Suid-Afrikaanse universiteite, naamlik die Universiteit van Pretoria (UP), die Universiteit van Wes-Kaapland (UWK) en die Universiteit van die Witwatersrand (UW) onder die loep neem, is onder andere bevind dat tot 20%

van voorgraadse studente aan die UP hulle studie staak en dat minder as 40% dit binne die minimum tyd wat toegelaat word, voltooi. In die tydperk tussen 2001 en 2004 het gemiddeld 18,5% van voorgraadse studente by die UWK nie met universiteit klaargemaak nie, terwyl die oorgrote meerderheid studente vier jaar of meer benodig om 'n driejaarkursus te deurloop. By die UW verwerf gemiddeld minder as die helfte van voorgraadse studente uiteindelik hulle kwalifikasie en minder as 45% gradueer op tyd. Volgens die Suid-Afrikaanse Instituut vir Rasseverhoudinge het slegs 14% van Suid-Afrikaanse studente wat in 2008 vir 'n driejaargraadkursus ingeskryf het aan die einde van 2010 hulle grade behaal (Jansen, 2013:5). Vir die betrokke tersiêre inrigtings verteenwoordig hierdie syfers groot geldelike verliese in die vorm van subsidies (Rademeyer, 2005), en vir die ouers van studente beteken dit verspilde fondse.

Die onvermoë van skole om leerders voldoende vir universitêre studie toe te rus, word as een van die hooforsake van lae deurvloeikoerse beskou (Van Dyk, 2011; Van Rooy & Coetzee-Van Rooy, 2015). As gevolg van die ontoereikendheid van die skoolkurrikulum en die Nasionale Senior Sertifikaat (NSS) het die oorgang van sekondêre na tersiêre studie 'n buitengewoon veeleisende sprong vir studente geword. Weens die “lae matriekstandaarde” word universiteite “oorval [...] met eerstejaars” wat “nie die mas kan opkom nie” (Nel, 2014:9), hoofsaaklik vanweë 'n gebrek aan akademiese geletterdheidsvaardighede, of so word algemeen aanvaar. Verskeie studies dui daarop dat swak akademiese prestasie met lae akademiese geletterdheidsvlakke in verband gebring kan word (Weideman & Van Rensburg, 2002; Van Dyk & Weideman, 2004; Van Dyk, 2011; Van Dyk & Van der Slik, 2012). Weideman en Van Rensburg (2002:158) ag leesvaardigheid, wat onder die begrip akademiese geletterdheid ressorteer, byvoorbeeld as 'n onontbeerlike voorvereiste vir akademiese sukses. Dit is, volgens Van Dyk (2010:3), vanselfsprekend dat studente wat taalmatig swak voorbereid is, of wie se akademiese lees- en skryfvermoë benede peil ontwikkel is, sal sukkel om met akademiese materiaal om te gaan en hulle eie materiaal opnuut te skep.

Vir die doeleindes van hierdie studie is dit relevant om daarop te wys dat probleme met akademiese taalvaardigheid nie slegs tot niemoedertaalsprekers van Engels beperk is nie. Hoewel Engels as onderrigtaal in die tersiêre omgewing oorheers en die getal studente vir wie dit nie 'n eerste taal is nie, ongetwyfeld aan die toeneem is (Van Dyk, 2010:7-9; Rambiritch, 2012:1-4), is dit ook so dat “problems with language proficiency [indeed] not only [affect] non-mother tongue users of academic language, but also first language speakers of English and Afrikaans” (Weideman & Van Rensburg, 2002:153). Selfs studente wat in hulle moedertaal studeer, ervaar die akademiese diskoers toenemend as 'n “vreemde taal” wat met moeite aangeleer en bemeester word (Gee, 1998; Van Dyk, 2010:5; Rambiritch, 2012:1; Webb, 2006:39).

In 'n poging om hierdie vraagstuk binne die hoër onderwys op te los, sonder belemmering van groter toeganklikheid en ooreenkomstig hierdie sektor se verbintenis tot gelykberegtiging (RHO, 2010:21-23), is daar landwyd by tersiêre inrigtings 'n verskeidenheid inisiatiewe in die lewe geroep om swak voorbereide studente in terme van hulle akademiese ontwikkeling te ondersteun. Één so 'n strategie is die aanbied van kursusse in akademiese geletterdheid, gewoonlik deur 'n afdeling soos die Eenheid vir Taalontwikkeling aan die Universiteit van die Vrystaat (UV), wat daarop gemik is om studente se akademiese taalvermoë te help verbeter. Die gebruik van taaltoetsing by universiteite is 'n direkte uitvloeisel van die behoefte om te bepaal watter studente potensiële risikostudente met lae akademiese geletterdheidsvlakke is wat waarskynlik by ondersteuningsmeganismes soos die voorgenoemde baat sal vind. Van Rooy en Coetzee-Van Rooy (2015:2) merk op dat “universities are challenged to use whatever information is available to predict students' chances of success, and also to determine the risk profile of entering students to decide on the best type of support to offer students in order to enhance their chances of success”.

Bachman en Purpura (2010:456) dui in hierdie verband aan dat taaltoetse onvermydelik bepaalde uitsprake of beslissings ten opsigte van die kandidate wat



dit aflê, tot gevolg het, en dat die tipe besluite wat op grond van hierdie toetse se uitslae geneem word, uiteenlopend van aard kan wees. Byvoorbeeld, waar sommige universiteite, soos pas vermeld, taaltoetse aanwend om te help bepaal watter ingeskrewe studente die gevaar loop om akademies te onderpresteer, gebruik ander dit om toelating tot die betrokke inrigting of tot 'n spesifieke fakulteit, departement of graadkursus te reguleer (sogenaamde hoëbelangtoetse). In die woorde van Bachman en Purpura (2010:465) “language tests serve as gate-keepers for some and door-openers for others”. Die sosiale impak wat taaltoetsing kan hê (McNamara & Roever, 2006) blyk duidelik uit hierdie stelling – 'n aspek van toetsontwerp wat betrekking het op die regulatiewe voorwaardes vir toetsontwerp (Weideman, 2017) wat in Hoofstuk 2 van hierdie studie onder meer aan die orde kom. Nietemin, al benut universiteite taaltoetse nie noodwendig op dieselfde manier nie, wil dit voorkom asof die assessering van akademiese taalvaardigheid 'n instelling geword het waarvan daar waarskynlik nie in die nabye toekoms afgesien sal word gegewe studente se lae geletterdheidsvlakke nie (Jansen, 2013:5; Nel, 2014:9). Trouens, volgens Maree (in Jansen, 2013:5), “benchmark tests are playing an ever-increasing role”.

Só byvoorbeeld word die Nasionale Normtoets (NNT) of National Benchmark Test (NBT) wat 'n toets van akademiese geletterdheid insluit, tans deur verskeie tersiêre instellings (onder andere die UV) vir die evaluering van eerstejaarstudente se akademiese gereedheid ingespan. Buiten hierdie toets is daar ook die Test of Academic Literacy Levels (TALL) wat aanvanklik deur die UP in samewerking met die Noordwes-Universiteit (NWU) en Universiteit Stellenbosch (US) ontwikkel is en vir die eerste keer in 2004 gebruik is. Hierdie vennootskap waarvan die UV op hierdie oomblik deel uitmaak, staan as die Inter-Institutional Centre for Language Development and Assessment (ICELDA) bekend. Heelparty plaaslike universiteite, soos die UP, die NWU en die US maak in hierdie stadium van TALL en/of sy Afrikaanse eweknie, te wete die Toets van Akademiese Geletterdheid (TAG), gebruik om te bepaal in watter mate eerstejaarstudente akademies geletterd is. TALL word soms ook deur studente

aan buitelandse universiteite, soos die Universiteit van Danang in Viëtnam (Le, 2011; Le, Du Plessis & Weideman, 2011) en die Universiteit van Namibië, afgeneem.

Dit sal egter 'n mistasting wees om die afleiding te maak dat gebrekkige akademiese geletterdheid uitsluitlik tot voorgraadse studente, veral eerstejaars, beperk is. Hoewel dit 'n uitvloeisel van laasgenoemde is, word daar vervolgens aangedui hoedat die bedenklike akademiese skryfvermoë van nagraadse studente (asook voornemende nagraadse studente) 'n kommerwekkende kwessie in eie reg is. Dié aangeleentheid is juis die beweegrede vir hierdie studie, wat die ontwerp en teoretiese regverdiging van 'n nagraadse toets van akademiese geletterdheid in Afrikaans omvat.

## 1.2 Gebrekkige akademiese geletterdheid: Die nagraadse nadraai

Vir Du Plessis (2012:3) is die moontlikheid van studente wat vir nagraadse studie inskryf sonder dat hulle gedurende hulle voorgraadse termyn 'n voldoende vlak van akademiese geletterdheid bereik het, 'n onrusbarende gedagte. Hierdie vrees blyk geregverdig te wees in die lig daarvan dat die akademiese eise wat aan nagraadse studente gestel word, omvangryk van aard is. Volgens Butler (2009:291) wat in sy doktorsale proefskrif (Butler, 2007) ondersoek ingestel het na die skryfprobleme van 'n groep nagraadse studente aan die UP “a number of postgraduate supervisors [...] confirmed that many postgraduate students still struggle with academic writing”. Die algemene gevoel onder die betrokke akademici wat hulle waarnemings met Butler gedeel het, was dat hulle nagraadse studente se akademiese geletterdheidsvlak van swak tot gemiddeld wissel. Dít terwyl die suksesvolle voltooiing van 'n nagraadse kwalifikasie, na hulle mening, moeilik haalbaar is indien 'n student nie oor 'n toereikende vlak van akademiese taalvaardigheid beskik nie. Sulke studente sukkel om akademies staande te bly, met die gevolg dat hulle dikwels nie hulle navorsingsprojekte klaarmaak nie, baie lank neem om dit af te handel en/of baie tyd en moeite in terme van

studieleiding verg (Butler, 2011). Butler (2009:296) wys daarop dat byna al die respondente die idee dat studente reeds akademies geletterd moet wees alvorens hulle toegang tot nagraadse studie verkry, ondersteun het. Ten einde sodanige maatreël in plek te stel, moet vooraf bepaal kan word of studente akademies geletterd is al dan nie. Hiervoor is 'n betroubare toets wat akkurate inligting oor kandidate se akademiese geletterdheidsvlak verskaf, nodig.

Butler (2007; 2009; 2011) se navorsing het die noodsaaklikheid van só 'n meetinstrument wat spesifiek vir nagraadse studente ontwerp is, blootgelê. Die Test of Academic Literacy for Postgraduate Students (TALPS) is in antwoord op hierdie behoefte deur die Eenheid vir Akademiese Geletterdheid aan die UP gedurende 2006/07 as deel van die Postgraduate academic literacy initiative ontwikkel, met fondse wat op 'n mededingende grondslag deur die RHO uit Finse bronne toegeken is. Hierdie toets word nie slegs plaaslik deur studente afgelê nie, maar is ook al deur die National University of Singapore (NUS) gebruik.

TALPS is op dieselfde konstruk van akademiese geletterdheid as TALL gebaseer. Hoofstuk 3 bied 'n uiteensetting van dié konstruk wat definieer watter vermoëns deur beide hierdie toetse gemeet word, asook van Patterson en Weideman (2013b) se onlangs uitgebreide weergawe daarvan, wat vir die doeleindes van hierdie studie as die toepaslike onderbou vir 'n Afrikaanse nagraadse toets van akademiese geletterdheid geïdentifiseer is. Oor die verwantskap tussen TALL en TALPS skryf Rambiritch (2012:35) die volgende:

The developers chose to base TALPS on the same construct as TALL. TALL was in many ways a sounding board for TALPS – the success of TALL was one of the most important factors that motivated the development of TALPS. Both these tests are designed to test the same thing: the academic literacy of students, the only difference being that one is directed at first-year students while the other is intended for postgraduate students.

Aangesien studente op nagraadse vlak by uitstek daartoe in staat behoort te wees om 'n goed geformuleerde akademiese argument op skrif te voer, verskil TALPS wel van TALL in dié opsig dat dit spesifiek klem op akademiese skryf plaas. Anders as TALL bevat dit 'n afdeling wat van kandidate vereis om 'n akademiese

teks te produseer<sup>1</sup>, asook 'n vraag oor teksredigering. Butler (2009:294) verduidelik die rasionaal agter die insluiting van hierdie twee subtoetse in TALPS soos volg:

In the development of TALPS the test developers have also considered the importance of testing students' productive writing ability specifically (in the production of an authentic academic text), as well as their editing ability. Although less emphasis is placed on writing in TALL (with specific reference to the writing of longer academic texts) because of practical constraints in administering the test, all the abilities contained in the definition of [academic literacy] are necessary preconditions for successful academic writing.

'n Verdere verskil tussen TALL en TALPS – wat as die motivering vir die aanpak van hierdie studie dien – is die feit dat TALL, as 'n voorgraadse toets van akademiese geletterdheid, oor 'n Afrikaanse ekwivalent, naamlik TAG, beskik, terwyl nagraadse studente tans geen ander keuse het as om in Engels geassesseer te word nie. Dat daar teenswoordig geen Afrikaanse alternatief vir TALPS beskikbaar is nie, word as 'n leemte in die gebied van taaltoetsing op tersiêre vlak beskou.

Gegewe die beleërde posisie van Afrikaans aan verskeie universiteite en vrese oor die toekoms van Afrikaans as universiteitstaal, kan die lewensvatbaarheid van 'n nagraadse toets van akademiese geletterdheid in Afrikaans met reg bevraagteken word. Om bloot bedreiging en politieke druk as motivering vir die nalaat van die bevordering van veeltaligheid, soos in die Grondwet in die vooruitsig gestel word, te gebruik, is egter nie voldoende rede om die poging om dit te bevorder, ook in die geval van ander inheemse tale, te laat vaar nie. In die volgende afdeling word daarom aangetoon dat daar in weerwil van die aanslag op Afrikaans as onderrigtaal 'n behoefte aan en ruimte vir só 'n toets bestaan.

---

1. 'n Verfyning van die nasienprotokol vir die skryfstuk in TALPS, asook van die diagnostiese inligting wat die toets bied, is intussen deur 'n Groningse navorser, Anna Pot, onderneem – 'n saak wat later in hierdie studie weer aan bod kom (Pot & Weideman, 2015).

### 1.3 Die behoefte aan 'n nagraadse toets van akademiese geletterdheid in Afrikaans

Bornman, Pauw en Potgieter (2014) wys daarop dat hoewel die Grondwet die bevordering van meertaligheid en taalgelykheid onderskryf, die onderwyslandskap in Suid-Afrika toenemend aan verengelsing onderwerp word. Hierdie outeurs (Bornman et al., 2014:598-600) en andere na wie hulle verwys, skryf die agteruitgang van moedertaalonderrig, veral wat universitêre onderrig in Afrikaans betref, aan 'n verskeidenheid faktore toe. In die eerste plek is enkelmedium Afrikaanse universiteite sedert die aanvang van die nuwe politieke bedeling in 1994 uitgefaseer, terwyl Afrikaanssprekendes tans, as 'n minderheidsgroep, beperkte politieke en ideologiese bedingingsmag in die openbare sfeer het. Verder word veeltaligheid ook deur swak beleidsimplementering by universiteite en 'n gebrek aan aanmoediging en ondersteuning van regeringsweë ondermyn. Boonop het Engels, onder die invloed van globalisering, status as “globale *lingua franca*” verwerf, en die bemeestering van Engels word gevolglik deur baie as 'n “globale vaardigheid” beskou. Dié siening dat Engels 'n wêreldtaal is, bring mee dat leerders, studente en ouers baie keer gedwonge voel om doelbewus 'n keuse téén moedertaalonderrig uit te oefen, desondanks aanduidings dat die ontwikkeling van akademiese taalvaardigheid in 'n tweede of addisionele taal problematies is. Webb (2006:39-40) merk in hierdie verband op dat “leerders nie skool toe [kom] met hierdie soort vaardigheid nie, nie eers in hulle moedertaal nie. [Akademiese taalvaardigheid] moet ontwikkel word, en dit neem etlike jare. In die geval waar die onderrigtaal 'n tweede/addisionele taal is, sal dit uiteraard nog langer neem”.

Soos wat die nie-optimale verwerwing van hoërordetaalvaardighede in 'n niemoedertaal studente uiteindelik opvoedkundig mag benadeel, kan die toetsing van dié vaardighede in 'n niemoedertaal logieserwys ook nie in hulle beste belang wees nie. In hierdie opsig sal 'n nagraadse toets van akademiese geletterdheid in Afrikaans vir Afrikaanssprekende studente 'n meer billike assesserings-

instrument as die huidige Engelse toets wees. Oor die regverdigheid van TALPS het Du Plessis (2012:123-124) in haar resepsiestudie van dié toets – waaraan 139 nagraadse studente van die Universiteit van die Vrystaat deelgeneem het – onder andere die volgende bevind:

On the matter of the language of testing, the results of the survey suggest that an Afrikaans alternative needs to be developed, even if only a small percentage of students elect not to do the test in English. This would bring the TALPS into alignment with the parallel-medium language policy of the university in terms of which students may choose to complete their studies in either English or Afrikaans and also provide a means to counteract the hegemony of English.

Hierdie bevinding wat deur die kommentaar van enkele studente, soos byvoorbeeld dat die toets makliker vir moedertaalsprekers van Engels is, gerugsteun word, is reeds vóór die verspreiding van die vraelys deur Du Plessis (2012:102) geantisipeer.

[S]tudents of minority or other groups whose mother-tongue languages were not English [...] could be expected to describe TALPS as unfair, since the test only recognises knowledge of the English language as a literacy tool and does not take into account levels of academic literacy attained in other languages.

Om ag te slaan op besware soos dié, is veral belangrik om onder andere twee redes. In die eerste plek verswak negatiewe persepsies omtrent 'n toets die gesigsgeldigheid daarvan. McNamara (2000:133) definieer gesigsgeldigheid onder meer as die mate waarin 'n toets aan die verwagtinge van sy gebruikers en diegene wat dit aflê, voldoen, en hoe aanvaarbaar hulle dit vind. 'n Afrikaanse gelyke vir TALPS sal, gegewe die uitgesproke behoefte daaraan, hoogs waarskynlik baie gunstig deur moedertaalsprekers van Afrikaans ontvang word, mits die ontwerp daarvan in geen ánder opsig gebrekkig blyk te wees nie. Terselfdertyd sal die gesigsgeldigheid van TALPS ook positief beïnvloed word, aangesien besware soos die bogenoemde deur die beskikbaarheid van 'n Afrikaanse alternatief vir TALPS uitgeskakel word.

Verder, omdat verantwoordelike toetsontwerp deur inagneming van onder meer etiese oorwegings, soos billikheid en regverdigheid, gekenmerk word, mag aantygings dat 'n toets onregverdiglik teen sekere individue diskrimineer, nie

geïgnoreer word nie.<sup>2</sup> Die siening van sommige Afrikaanse studente dat TALPS moedertaalsprekers van Engels bevoordeel, is egter geensins te wyte aan swak of onverantwoordelike toetsontwerp nie. Die etiese dilemma spruit eerder voort uit die feit dat Afrikaanse studente vanweë ’n gebrek aan die beskikbaarheid van ’n toets in hulle moedertaal geen ander keuse het as om in Engels geassesseer te word nie. Die beskikbaarstelling van ’n Afrikaanse toets aan universiteite sal vanuit ’n etiese oogpunt beskou, “care and concern for those whom [it] measures” (Weideman, 2011:111) demonstreer. Hoewel toetsing van ’n ander aard hier ter sprake is, bevestig ’n insident in 2013 by die Universiteit van Suid-Afrika (Unisa) – wat in daardie stadium nog afstandsonderrig in Engels sowel as Afrikaans aangebied het<sup>3</sup> – dat studente wat in ’n gekose taal studeer, maar dan in ’n ander geassesseer word, verontreg mag voel. Rademeyer (2013:13) berig dat Afrikaanse studente by hierdie instansie benadeel gevoel het toe slegs Engelse vraestelle tydens ’n eksamenleentheid aan hulle beskikbaar was.

Uit ’n regs oogpunt<sup>4</sup> sal ’n Afrikaanse eweknie van TALPS daartoe bydra dat nagraadse taaltoetsing in Suid-Afrika in ooreenstemming met die grondwetlike doelstelling van veeltaligheid gebring word. ICELDA, onder wie se vaandel TALPS ontwikkel is, stel taaldiversiteit immers hoog op prys. Hiervan getuig die volgende uitspraak op die ICELDA-webwerf oor die stigting van hierdie vennootskap in 2004: “It is no coincidence that [the] founding partners are universities that are studiously multilingual: such institutions have a heightened and urgent awareness of language” (ICELDA, 2015). Dit is egter geen geheim nie dat verskeie tradisioneel Afrikaanse universiteite, waaronder ook stigtingsvennote van ICELDA, se verbintenis tot veeltaligheid in die laaste jare merkbaar agteruitgegaan het. Buiten die NWU wat steeds “’n funksionele en meertalige taalbeleid” (Noordwes-Universiteit, 2016) volg met voorgraadse en

---

2. In sy teoretiese raamwerk vir die ontwerp van toegepaste linguïstiese artefakte identifiseer Weideman (2017) veertien aspekte of dimensies wat die verantwoordelike ontwerp van taaltoetse ten grondslag lê, waarvan etiese bewustheid een is. Hierdie raamwerk kom in Hoofstuk 2 onder bespreking.

3. Unisa se raad het op 28 April 2016 die besluit geneem dat hierdie instelling se taalbeleid met ingang van 2017 verander na nét Engels (Gous, 2016).

4. Buiten die etiese, herken Weideman (2017) ook die juridiese as een van die veertien dimensies waarop verantwoordelike taaltoetsontwerp berus.

nagraadse programme in Engels of Afrikaans, het die UV en UP in 2016 aangekondig dat met ingang van 2017 Engels die primêre voertaal van beide hierdie instellings word, met uitsluiting van enkele fakulteite wat steeds 'n keuse oor die taal van onderrig sal hê (Van Rooyen, 2016; Strydom, 2016). Ook die US het sy taalbeleid in 2016 hersien en hoewel sommige kritici van mening is dat “dit moeilik [is] om te sien hoe dit in die praktyk tot ware meertaligheid sal lei” (Die Afrikaanse Taalraad, 2016), verklaar die beleid nietemin dat “geen studente van die US se akademiese aanbod uitgesluit sal word op grond daarvan of hulle Afrikaans of Engels magtig is nie” (Universiteit Stellenbosch, 2016a). Wat nagraadse onderrig en leer betref, “kan enige taal gebruik word, met dien verstande dat al die betrokke studente vaardig genoeg in daardie taal is” (Universiteit Stellenbosch, 2016b), 'n vereiste wat ironies genoeg stukrag verleen aan die nood van meertalige akademiese geletterdheidstoetse op nagraadse vlak by hierdie instelling.

Dat Afrikaans dus in die huidige tersiêre bestel swakker daaraan toe is, is onteenseglik, maar om aan te voer dat Afrikaans in die hoëronderwysomgewing naby aan sy einde gekom het, sou onverskillig en ook onakkuraat wees. Hoewel beperk, is daar tans, soos aangetoon, steeds 'n meertalige aanbod beskikbaar by enkele hoofstroomuniversiteite en bestaan daar ook taalinisiatiewe soos ICELDA wat, ongeag die heersende toestande, 'n durende verbintenis tot die bevordering van veeltaligheid op tersiêre vlak handhaaf en uitbou (Weideman, 2014). Dit is ook nie te ontken nie dat daar uit studentegeledere steeds 'n lewendige behoefte bestaan om gevorderde grade in Afrikaans te doen. Volgens dr. Dioné Prinsloo, die uitvoerende hoof van die Suid-Afrikaanse Akademie vir Wetenskap en Kuns is die Akademie se beurskantoor in 2016 “oorval met aansoeke van talle studente wat hulle meesters- en doktorsgrade in Afrikaans wil onderneem” (Maroela Media, 2016). Boonop blyk dit toenemend dat daar ontluikende groeimoontlikhede vir Afrikaans is by privaat hoëronderwysinstansies soos Akademia, die privaat Afrikaanse universiteit van die Solidariteit Beweging (Gericke, 2016), asook by AROS, die privaat



onderwys-opleidinginstelling in Pretoria, waar sommige van die loodstoetse vir hierdie studie gedoen is.

In die lig van die voorafgaande is dit duidelik dat daar genoeg gronde vir die ontwerp en gepaardgaande teoretiese regverdiging van 'n nagraadse toets van akademiese geletterdheid in Afrikaans is. Afrikaans kan, in die woorde van Lina Spies (2014:17) “nie behou [word] deur defensiewe optredes en deur 'n futiele verzet teen die internasionale relevansie van Engels nie”, maar wel deur “produktiewe optredes” waarvan die skep van 'n Afrikaanse nagraadse taaltoets seer seker een is.

## **2. Doel van die studie**

Die rasionaal vir hierdie studie wat in die vorige afdeling in besonderhede verduidelik is, omvat in wese 'n versameling verbandhoudende probleme waarvoor 'n geskikte oplossing gevind moet word.<sup>5</sup> Hierdie probleme kan kortliks soos volg saamgevat word:

- Deurvloeiakoerse en uitvalsifers by tersiêre onderwysinrigtings in Suid-Afrika is onbevredigend wat kleiner subsidies vanaf owerheidsweë, asook huishoudelike of persoonlike geldelike verlies, tot gevolg het.
- Weens lae akademiese geletterdheidsvlakke is akademiese sukses vir veral voorgraadse studente moeilik bereikbaar, moedertaalsprekers van Afrikaans inkluis.
- Die nasleep hiervan is dat die akademiese skryfvermoë van nagraadse studente ook onder verdenking is, wat onbelemmerde vordering met studie kniehalter.
- Alhoewel 'n nagraadse toets van akademiese geletterdheid (TALPS) reeds ontwikkel is ten einde te bepaal in watter mate voornemende nagraadse

---

5. Hierdie waarneming strook met die hedendaagse opvatting dat die toegepaste linguistiek, die vakgebied waarbinne hierdie studie val, 'n dissipline is wat oplossings vir talige probleme bied. Die siening dat die toegepaste linguistiek 'n dissipline van die ontwerp van sulke oplossings is, word in Hoofstuk 2 behandel.

studente akademies geletterd is, is hierdie toets slegs in Engels beskikbaar.

- Die tekort aan 'n gelykwaardige Afrikaanse weergawe van TALPS lei tot aanklagte van onbillikheid deur studente wat in Afrikaans studeer, asook tot die miskenning van veeltaligheid.

Hoewel die eersgenoemde twee uitdagings hierbo met die res verband hou, het die doel van hierdie studie regstreeks op die laaste drie probleme betrekking. Die oorkoepelende, rigtinggewende oogmerk van hierdie ondersoek is om in die behoefte aan 'n nagraadse toets van akademiese geletterdheid in Afrikaans te voorsien. Die bereiking hiervan hang ten nouste saam met die verwesenliking van vier verdere, meer spesifieke doelwitte wat hieronder uiteengesit word.

In die eerste plek het die studie ten doel om uitvoering te gee aan die ontwerp, loodsing en verfyning van 'n Afrikaanse nagraadse toets van akademiese geletterdheid – aktiwiteite wat tipiese momente in die proses van taaltoetsontwikkeling verteenwoordig. Volgens Bachman en Palmer (1996:86-93, raadpleeg ook Read, 2010:290-293) begin die ontwerpproses by die artikulasie van 'n toetskonstruk wat 'n omskrywing van die vermoëns wat deur die bepaalde toets gemeet word, verskaf. Die volgende stap behels die operasionalisering van hierdie konstruk by wyse van die ontwikkeling van toetsspesifikasies en toepaslike taaktipes. Hierdie taaktipes word vervolgens aan loodstoetsing onderwerp ten einde die produktiewe items te identifiseer, die onproduktiewe items aan te pas of te skrap en die tegniese kwaliteit (d.i. onder meer die betroubaarheid en geldigheid) van die toets te evalueer. Hoofstuk 3, 4 en 5 doen verslag oor hoe hierdie ontwikkelingsproses vir 'n Afrikaanse nagraadse toets van akademiese geletterdheid verloop het, en staan sodoende in diens van die teoretiese regverdiging van hierdie toets, wat 'n verdere doelwit van dié studie is.

Van Dyk (2010:46) wys daarop dat 'n taaltoets (wat ooreenkomstig die toegepaste linguistiek as 'n oplossing vir 'n taalprobleem beskou kan word – kyk Hoofstuk 2), “nie bloot ontwerp en geïmplementeer [kan] word [...] sonder 'n deeglike regverdiging daarvan in terme van wetenskaplike teorie en analise [nie], en ook nie sonder om dit tegnies verantwoordbaar en met integriteit te doen nie”. Hierdie studie stel dit daarom vernaamlik ten doel om aan te toon hoedat 'n teoreties-gefundeerde, wetenskaplik-verantwoordbare benadering tot toetsontwerp die ontwikkeling van die betrokke Afrikaanse nagraadse toets van akademiese geletterdheid ten grondslag lê. Die teoretiese en eties regverdigbare begroning vir die ontwerp van hierdie toets, geskied aan die hand van Weideman (2017) se raamwerk van ontwerpbeginsels vir die toegepaste linguistiek. Sy raamwerk dien as die teoreties-filosofiese vertrekpunt vir hierdie studie en word in Hoofstuk 2 in besonderhede bespreek. Weideman (2017) bring verantwoordelike toetsontwerp eensyds met die inagneming van intrinsieke, konvensionele voorwaardes soos betroubaarheid en geldigheid (die sogenaamde konstitutiewe voorwaardes vir toetsontwerp) in verband, en andersyds met die verrekening van sosiale aspekte soos deursigtigheid, verantwoordbaarheid en regverdigheid (die sogenaamde regulatiewe voorwaardes). Vir hierdie outeur kan integere toetsontwerp slegs realiseer wanneer daar 'n gesonde balans tussen die voorgenoemde konstitutiewe en regulatiewe dimensies gehandhaaf word. Volgens Weideman (2011:111) “[t]ests derive their integrity from their theoretical defensibility, as well as from their social transparency or public accountability. They must not only be consistent, but their scores must also be responsibly interpreted”. Die rede waarom daar spesifiek op hierdie raamwerk as teoretiese basis besluit is, is dat dit 'n verruimde, gebalanseerde perspektief omtrent die verantwoordbare ontwerp van taaltoetse bied (Van Dyk, 2013). Soos uit Hoofstuk 2 duidelik sal word, verteenwoordig Weideman se benadering tot verantwoordelike toetsontwerp 'n alternatiewe, meer gematigde beskouing wat sigself tussen twee uiterste paradigmas, naamlik 'n tradisionele en die hedendaagse ortodoksie, posisioneer.

Die derde doelwit van hierdie studie hou verband met die konstruk van akademiese geletterdheid waarop die beoogde nuwe toets gebaseer word. Hoewel die konstruk wat vir die ontwerp van TALL, TAG en TALPS gebruik is, 'n goed gevestigde, teoreties verdedigbare definisie van akademiese geletterdheid omvat, het Patterson en Weideman (2013b) dit goedgevind om hierdie konstruk in heroënskou te neem, aangesien “no critical examination of [this] construct which is now more than a decade old has so far been undertaken”. In hulle hersiene weergawe van dié konstruk gee hulle uitdrukking aan wat hulle beskou as 'n “[improved] [...] understanding of what makes academic discourse unique (Patterson & Weideman, 2013a:107). Hierdie outeurs is van mening dat 'n definisie van akademiese geletterdheid ontstaan uit 'n idee van wat akademiese diskoers eiesoortig of anders as ander diskoerstipes maak. 'n Beter begrip van wat akademiese diskoers tipies maak, sal daarom 'n meer trefseker omskrywing van akademiese geletterdheid tot gevolg hê. Ná 'n kritiese oorsig van verskeie beskouings omtrent die aard van akademiese taal, kom Patterson en Weideman (2013a & 2013b) tot die gevolgtrekking dat die eiesoortigheid van akademiese diskoers in die logiese en analitiese taal waardeur dit gekenmerk word, te vinde is. Hulle motiveer hulle daaruit volgende aanpassing van die oorspronklike konstruk wat vir die ontwikkeling van TALL, TAG en TALPS gebruik is, soos volg (2013b:137):

If the typicality of academic discourse is derived from the (unique) distinction-making activity associated with the analytical or logical mode of experience, then that aspect needs to take precedence. In [the earlier] definition, distinction-making and analytical or logical thinking are indeed a component, yet it is not foregrounded as the most important aspect of academic literacy.

Patterson en Weideman (2013b:139-140) skuif dus in hulle bygewerkte definisie van akademiese geletterdheid – wat volledig in Hoofstuk 3 weergegee en krities bespreek word – die vermoë om onderskeidings op grond van logiese en analitiese denke te maak – op die voorgrond. Hierdie hersiening maak geensins 'n algehele wegbeweeg van die huidige konstruk uit nie, maar is eerder 'n verfyning daarvan: “[i]t suggests on a number of points ways in which one might add to the current definition of academic literacy” (Patterson & Weideman, 2013b:126). Wysigings aan 'n bestaande toetskonstruk impliseer egter

gepaardgaande aanpassings ten opsigte van die operasionalisering daarvan. Patterson en Weideman (2013b:140-145) maak gevolglik, in ooreenstemming met hulle uitgebreide definisie van akademiese geletterdheid, verskeie voorstelle wat wysigings aan bestaande subtoetse en taaktipes, asook die skep van en eksperimentering met nuwes betref.

Vanweë die innoverende karakter van hierdie studie wat beoog om aan 'n Afrikaanse nagraadse toets van akademiese geletterdheid (én die teoretiese regverdiging daarvan) gestalte te gee, word dit as toepaslik geag om die ontwerp daarvan met Patterson en Weideman se vernude definisie van akademiese geletterdheid te belyn. Ten einde hierdie belyning te implementeer, moet daar soos wat Patterson en Weideman aan die hand doen, gewysigde en/of nuwe subtoetse en/of taaktipes geskep, en die produktiwiteit daarvan deur middel van loodstoetsing, geëvalueer word. Laasgenoemde maak die derde doelwit van dié studie uit. Daar word beplan om enkele taaktipes en items wat nog nie vantevore in TALL, TAG of TALPS verskyn het nie, en wat veral logiese en/of analitiese denke betrek, by die loodswaegawe van die voorgestelde Afrikaanse toets in te sluit.

Watter items uiteindelik in die finale waegawe van die toets opgeneem gaan word, sal deur statistiese analise van die toetsresultate na afloop van die loodsfase(s) bepaal word. Vir 'n item om aanvaar te word, moet dit, onder andere, 'n diskriminasiewaarde (d.i. hoe goed 'n item tussen sterk en swak kandidate onderskei) van meer as 0.15 en 'n fasiliteitswaarde ('n aanduiding van die moeilikheidsgraad van 'n item) van tussen 0.15 en 0.84 hê. In Hoofstuk 4 word in besonderhede aangetoon wat die geldende kriteria en parameters vir die kwantitatiewe metings van itemproduktiwiteit in hierdie studie is. In Hoofstuk 5 word aangedui of die eksperimentering met gewysigde en/of nuwe subtoetse en/of taaktipes enigsins geregverdig was.

Die vierde en laaste doelwit het te doen met die formaat waarin die beoogde toets uiteindelik sy gestalte gaan kry. Hierdie studie stel dit as mikpunt om genoeg toetsitems tydens die ontwerpfasie te genereer ten einde toekomstige eksperimentering met 'n meerfasige toetsmodel, spesifiek 'n “three-tiered test model” of driefaseontwerp, moontlik te maak. Daar word dus onderneem om 'n eerste tree te gee tot die moontlike implementering van 'n meerfasige, in plaas van eenfasige, akademiese geletterdheidstoets, soos wat tans met bv. TALPS die geval is. 'n Uiteensetting van die redes vir die operasionalisering van die beoogde toets as 'n potensiële driefasetoets kan in Hoofstuk 3 gevind word.

### **3. Navorsingsmetodologie**

Die ondersoekgebied wat hier ter sprake is, is taaltoetsing wat as 'n subdissipline van die toegepaste linguïstiek gekategoriseer word (kyk Hoofstuk 2). As navorsingswerk behoort hierdie studie derhalwe tot die terrein van die toegepaste linguïstiek. Sorg is daarom gedra dat die navorsingsmetode waarvolgens daar in hierdie studie te werk gegaan word, doelmatig raakpunte het en in verskeie opsigte saamval met beskrywings van hoe wetenskapsbeoefening in die toegepaste linguïstiek benader kan word, gegewe die verskeidenheid klassifikasies en werkwyses wat binne hierdie veld aanneemlik en toepasbaar is.

Brown (2004:478) definieer navorsing in die toegepaste linguïstiek as 'n “systematic and principled inquiry in applied linguistics” – 'n omskrywing wat weerklank vind in die navorsingsontwerp van hierdie studie. Nie alleen word daar onderneem om die ontwikkeling van die betrokke toets sistematies aan die hand van spesifieke stappe uit te voer nie, maar ook teoreties en eties beginselvas volgens 'n raamwerk van voorwaardes vir verantwoordelike toetsontwerp.

Die navorsingsmetode behels verder, ooreenkomstig Brown (2004:478-479) se kategorisering van navorsingstipes in die toegepaste linguïstiek, 'n kombinasie van sekondêre en primêre navorsing. Brown (2004:478) verduidelik dat

sekondêre navorsing uit die ondersoek en skryfwerk van ander navorsers ontstaan, terwyl primêre navorsing spruit uit oorspronklike data van die een of ander soort. Vir die teoretiese regverdiging van die ontwerp van die onderhawige geletterdheidstoets word op beide hierdie tipes navorsing gesteun. Om te begin word daar 'n gefokusde literatuurstudie uitgevoer waarin tersaaklike standpunte oor taaltoetsing wat uit verskeie sekondêre bronne verkry is, byeengebring word. Daar word byvoorbeeld aangedui hoedat taaltoetsing 'n onderdeel van die toegepaste linguïstiek is en hoedat verskillende paradigmas binne die toegepaste linguïstiek 'n invloed op die ontwikkelingsgeskiedenis van taaltoetsing gehad het. Daar word ook deur middel van 'n oorsig oor enersyds die konvensionele, en andersyds die meer eietydse opvattinge omtrent die bereiking van geldigheid in toetsontwerp, aangetoon hoedat deskundiges op die gebied van taaltoetsing “[have] struggled with issues of continuity, integrity and validation” (Weideman, 2006:71). Weideman se raamwerk van ontwerpbeginsels wat in reaksie op hierdie stryd in taaltoetsgeledere geskep is, word as 'n alternatiewe benadering tot integere toetsontwerp aangebied, en vorm die teoretiese uitgangspunt van hierdie studie. Weideman (2011:111) is trouens van mening dat “an approach to the design and development of language tests that takes as its point of departure a comprehensive set of design principles [...] is more likely to provide us with a responsible, defensible instrument”.

Verder, vanaf Hoofstuk 3 waar die fokus verskuif van taaltoetsing in die algemeen na spesifiek akademiese geletterdheidstoetsing, word daar deurlopend na literatuur verwys wat wesenlik betrekking het op die praktyk van (of die verskillende stappe in) die ontwerp van die onderhawige akademiese geletterdheidstoets. Dit behels inligting wat verband hou met die ontwikkeling en operasionalisering van 'n toetskonstruk, die loodsing van toetsitems en die daaropvolgende statistiese analise van toetsresultate, asook die verfyning van toetsitems.

Aanvullend by die sekondêre navorsing, betrek hierdie studie ook primêre navorsing. Primêre navorsing in die toegepaste linguistiek kan, volgens Brown (2004:479), kwantitatief en/of kwalitatief van aard wees, hoewel 'n duidelik afgebakende onderskeid nie altyd te tref is nie. Brown (2004:488) wys daarop dat onderskeiding tussen 'n kwantitatiewe en/of kwalitatiewe navorsingsbenadering in baie gevalle eerder 'n kwessie van graadverskil op 'n interaktiewe kontinuum as 'n skerp omlynde tweedeling is. Dit blyk dan ook die geval in hierdie studie te wees, aangesien beide kwantitatiewe en kwalitatiewe tegnieke en/of data vir die teoretiese verdediging van die onderhawige toets aangewend word. Kwantitatiewe ondersoekmetodes, soos empiriese proefneming (in die vorm van loodstoetsing) en statistiese ontleding van die loodstoetsresultate, word byvoorbeeld gebruik om bewyse ten opsigte van die betroubaarheid en geldigheid van die toets te genereer. Tog, omdat verantwoordelike toetsontwerp volgens huidige konvensie met meer as “dit wat gemeet en bewys kan word” (Van Dyk, 2013:160) in verband gebring word<sup>6</sup>, is dit vanselfsprekend dat kwalitatiewe tegnieke, soos beskrywing en interpretasie, deel van die navorsingsontwerp van 'n studie soos hierdie sal uitmaak. Trouens, volgens Hughes (2003:65), is beide kwantitatiewe en kwalitatiewe analise 'n kenmerkende gegewe van die metodologie van taaltoetsontwerp:

The first [kind of analysis] – statistical. This will reveal qualities (such as reliability) of the test as a whole and of individual items (for example, how difficult they are, how well they discriminate between stronger and weaker candidates). The second kind of analysis is qualitative. Responses should be examined in order to discover misinterpretations, unanticipated but possibly correct responses, and any other indicators of faulty items.

Verder, indien Brown (2004:491-496) se kontinuum van standarde vir bevredigende navorsing in die toegepaste linguistiek as maatstaf gebruik word, blyk dit duidelik hoedat hierdie studie elemente van beide kwantitatiewe en kwalitatiewe navorsing inkorporeer. Volgens Brown (2004:492) heg navorsers wat kwantitatiewe ondersoeke doen gewoonlik waarde aan konsepte soos betroubaarheid, herhaalbaarheid, geldigheid en veralgemeenbaarheid, terwyl

---

6. In Hoofstuk 2 word aangedui hoedat die tradisionele benadering tot geldige toetsontwerp van die hedendaagse een verskil.



beginsels soos verantwoordbaarheid en deursigtigheid veral in kwalitatiewe navorsing van belang is. In eersgenoemde val die klem op empiriese feite wat in syfers uitgedruk word, waar laasgenoemde op die idee berus dat “our observations need not only interpretation, but also sometimes political analysis and action” (Weideman, 2013:4480). In hierdie studie word die ontwerp van ’n Afrikaanse nagraadse toets van akademiese geletterdheid juis in terme van eenderse standaard as die voorgenoemde geregverdig – vergelyk in hierdie verband Weideman (2017) se konstitutiewe en regulatiewe voorwaardes vir toetsontwerp. Erkenning moet hier aan Van Dyk (2010:22) gegee word wat reeds in sy doktrale proefskrif opmerk hoedat die werk van Weideman (2006) en Brown (2004) by mekaar aansluiting vind.

#### **4. Waarde van die studie**

Hierdie studie lewer op verskeie maniere ’n bydrae tot die gebied van akademiese geletterdheidstoetsing in die Suid-Afrikaanse hoëronderwysomgewing. In die eerste plek sal die beskikbaarstelling van ’n Afrikaanse nagraadse akademiese geletterdheidstoets dubbel- en parallelmediumuniversiteite wat wel Afrikaans as voertaal gebruik, in staat stel om nagraadse studente wat kies om in Afrikaans te studeer se akademiese geletterdheid in hulle moedertaal te assesser. Dit kom neer op meer billike akademiese geletterdheidstoetsing wat boonop veeltaligheid help bevorder. Die verwagting is verder dat hierdie toets sal bydra tot ’n afname in die getal Afrikaanse studente wat weens swak akademiese taalvaardigheid op nagraadse vlak misluk. Afhangende van die beleid wat ’n bepaalde fakulteit of departement volg, kan studente wie se toetsuitslag benede peil is, byvoorbeeld verplig word om relevante ondersteuningsmodules of nagraadse skryfwerkswinkels te deurloop alvorens hulle met nagraadse studie mag begin. Du Plessis (2012:4-5) waarsku egter dat gebruikers nie onrealistiese verwagtinge van taaltoetse moet koester nie. Sy (2012:4-5) stel dit onomwonde dat “academic literacy tests are not psychometric tests designed to test potential, but rather instruments that reflect the current literacy level of a candidate”. Verskeie

gesaghebbendes op die terrein van taaltoetsing is dit daarom met mekaar eens dat wanneer daar besluite oor 'n kandidaat se toekoms geneem word geletterdheidstoetse verkieslik nie alleen nie, maar wel in samehang met ander aanwysers van akademiese sukses gebruik moet word (Du Plessis, 2012:5). Om hierdie rede geld die voorskrif dat akademiese departemente, aan 'n universiteit soos die UV, nie 'n voornemende nagraadse student slegs op grond van sy/haar TALPS-uitslag van studie mag uitsluit nie. Vir besluite omtrent toelating is die amptelik-geldende aanbeveling by hierdie universiteit dat die gewig van TALPS-toetsresultate nie die internasionaal aanvaarde perk vir die gewig wat aan taaltoetse toegeken word, naamlik tussen 10% en 20%, mag oorskry nie. Gewoonlik word 'n gewig van ongeveer 60% aan vroeëre akademiese prestasie toegeken, terwyl taalvermoë saam met verskeie ander kriteria tot die oorblywende 40% bydra (Hay, 2010).

Hierdie studie gee egter nie net aan 'n akademiese geletterdheidstoets as sodanig gestalte nie, maar ook aan die teoretiese regverdiging daarvan. Trouens, die wesenswaarde van hierdie studie is daarin geleë dat dit, ter ondersteuning van die integriteit en teoretiese verdedigbaarheid van die onderhawige toets, bewyse aanbied wat gegrond is op 'n raamwerk van ontwerpbeginsels vir verantwoordelike toetsontwerp. Dit behels terselfdertyd ook 'n aanduiding van hoe robuus hierdie raamwerk is, en hoe nuttig dit moontlik in die ontwerp van ander toegepaste linguistiese artefakte mag wees.

Ten laaste versit hierdie studie 'n baken in die opsig dat daar in die toetsontwikkelingsproses met enkele aangepaste en/of nuwe toetsitems wat tot nog toe nie in TALL, TAG of TALPS gebruik is nie, geëksperimenteer word. Omdat die toetsontwerp op 'n vernude konstruk van akademiese geletterdheid berus (vergelyk Patterson & Weideman, 2013b), is vernuwing in terme van die operasionalisering van hierdie verruimde konstruk 'n noodwendige gevolg. Meer as enigeen van sy voorgangers, plaas die hersiene konstruk die vermoë om

onderskeidings op grond van logiese en analitiese denke te maak eksplisiet op die voorgrond.

## **5. Uiteensetting van die studie**

Hieronder volg 'n kort uiteensetting van die inhoud van elke hoofstuk in hierdie studie.

Hoofstuk 1 gee inligting omtrent die agtergrond van hierdie studie en situeer dit binne die breër problematiek van akademiese ongereedheid en tersiêre onderprestering in die Suid-Afrikaanse hoërondewyskonteks. Daar word spesifiek aangedui wat die beweegrede vir, die doel van en die navorsingsbenadering tot hierdie studie is. Die waarde van die navorsing word ook uitgelig.

Hoofstuk 2 maak die kerngedeelte van die literatuurstudie uit en verskaf 'n grondige teoretiese basis aan hierdie studie. Dit ontsluit taaltoetsing as 'n subdissipline van die toegepaste linguistiek wat deur verskeie teoretiese strominge binne hierdie veld beïnvloed is. Wat die bestrede kwessie van integriteit in toetsing betref, bied hierdie hoofstuk 'n kritiese oorsig oor drie moontlike benaderings tot geldige toetsontwerp waarvan die genoemde raamwerk van Weideman (2017) wat hierdie studie teoreties ten grondslag lê, een is.

Met Hoofstuk 3 neem die dokumentering van die ontwikkelingsverloop van die Toets van Akademiese Geletterdheid vir Nagraadse Studente (waarna voortaan as TAGNaS verwys sal word) 'n aanvang. Hierdie hoofstuk bespreek die ontwikkeling en onlangse hersiening van die konstruk van akademiese geletterdheid waarop die onderhawige toets gebaseer word, asook die operasionalisering daarvan in die vorm van nuwe toetsspesifikasies en toepaslike taaktipes. Dit dui breedvoerig aan watter subtoetse vir TAGNaS beplan word, en

hoe hierdie subtoetse, waarvan sommige gewysigde of nuwe subtoetse is, aansluiting vind by die analitiese karakter van akademiese diskoers. Verder word daar ook vir 'n verskeidenheid ander ontwerpaspekte van die beplande assesseringsinstrument in hierdie hoofstuk motivering verskaf.

Hoofstuk 4 doen verslag van die resultate van die eerste loodstoetsing van TAGNaS en die daaropvolgende verfyning/seleksie van die onderskeie taaktipes/subtoetse in terme van hulle produktiwiteit. Daar word aan die hand van verskeie statistiese tegnieke ondersoek ingestel na die betroubaarheid en geldigheid van die toets as geheel, sowel as die produktiwiteit van die individuele toetsitems.

Hoofstuk 5 rapporteer die resultate van die tweede loodstoetsing van TAGNaS, hierdie keer die verfynde weergawe daarvan. Dit bied ook 'n evaluasie van die produktiwiteit van die gewysigde en/of nuwe subtoetse en/of taaktipes wat in hierdie studie op die proef gestel word. Met verwysing na statistiese analises van die toetsdata word aangetoon of die ontwerp en insluiting van dié betrokke items geregverdig was al dan nie.

Hoofstuk 6 bevat 'n gevolgtrekking waarin die verloop en bevindinge van hierdie studie saamgevat word. Die beperkinge van die studie word ook bespreek en aanbevelings vir verdere navorsing word gemaak.

## Hoofstuk 2

### Verantwoordelike toetsontwerp: 'n Teoretiese vertrekpunt

#### 1. Inleiding

Met Hoofstuk 1 as agtergrond hervat hiërdie hoofstuk die teoretiese begroning vir die ontwerp van die onderhawige nagraadse toets van akademiese geletterdheid, en dit gee 'n uitvoerige beskrywing van die teoretiese konteks waarbinne hierdie studie sigself posisioneer. In gevalle waar verantwoordelike ontwerp hoog op prys gestel word – soos in die gegewe een – is só 'n kontekstualisering van die uiterste belang. Weideman (2017:1) is van mening dat “[w]e cannot as applied linguists propose responsibly designed solutions to language problems if we remain ignorant of our theoretical foundations”. Tog, gegewe die dikwels dringende aard van toegepaste linguistiese werk – wat, soos vervolgens aangetoon word, die oplos van taalverwante sosiale probleme behels – laat toegepaste taalkundiges baiekeer na om behoorlik teoreties te besin omtrent dit wat hulle in die praktyk doen (Weideman, 2006:71). Omdat taaltoetsing (in hierdie geval spesifiek akademiese geletterdheidstoetsing) 'n praktyk is wat deel is van die toegepaste linguistiek, sal 'n teoretiese besinning oor eersgenoemde noodwendig refleksie ten opsigte van laasgenoemde insluit.

Om te begin, word die aard en historiese ontwikkeling van die toegepaste linguistiek as vakdissipline ondersoek, waarna 'n alternatiewe karakterisering daarvan as 'n dissipline van ontwerp verskaf word. Die argument sal gevoer word dat, alhoewel daar verskeie ander karakteriserings van die toegepaste taalkunde moontlik is, hierdie definisie besonder produktief aangewend kan word ten opsigte van taaltoetsing. Soos later duidelik sal blyk, vervul die nosie dat die toegepaste linguistiek 'n dissipline van ontwerp is, derhalwe 'n leidinggewende funksie binne die teoretiese raamwerk wat die toetsontwikkelingsproses in hierdie studie ten grondslag lê. Daarna kom taaltoetsing as 'n subdissipline van die toegepaste linguistiek onder die loep. 'n Verduideliking van die vernaamste

denkkrigtings binne hierdie veld word gegee, waarop toelgting omtrent die bestrede kwessie van geldigheid in toetsontwerp volg. Met die voorgenoemde bespreking as verwysingspunt, word die teoretiese raamwerk van ontwerpvoorwaardes waarop hierdie studie gefundeer is, ten slotte uiteengesit.

## **2. Die toegepaste linguistiek as 'n dissipline van ontwerp**

### 2.1 'n Definisie van die toegepaste linguistiek: Moderne en postmoderne opvattinge

Om die toegepaste linguistiek te definieer, behels 'n poging om die aard van hierdie veld te verduidelik. Beide Simpson (2011:2) en Wei (2011:8) bestempel Christopher Brumfit se omskrywing van die toegepaste linguistiek as die mees algemeen aanvaarde definisie van hierdie dissipline. Hy beskryf die toegepaste linguistiek as “the theoretical and empirical investigation of real-world problems in which language is a central issue” (Brumfit, 1995:27). Simpson (2011:2) wys daarop dat Brumfit se definisie om veral twee redes nuttig is. In die eerste plek is dit breed genoeg om die groot verskeidenheid ondersoekgebiede wat tans onder die toegepaste linguistiek ressorteer, te akkommodeer en, verder, onderskei dit die toegepaste linguistiek van ander verwante terreine deur dit spesifiek probleemgerig te maak.

While language is, of course, fundamental to human life, and surrounds us, the problem orientation helps to delimit the field. That is, the motivation for applied linguistics lies not with an interest in autonomous or idealized language, as with understandings of linguistics which deal in linguistic universals: applied linguistics data is typically collected empirically in contexts of use. Nor is its concern with the entirety of ‘language in use’. It is demarcated by its interest in how language is implicated in real-world decision-making. (Simpson, 2011:2)

Trouens, die probleemgerigheid van die toegepaste linguistiek neem 'n belangrike plek in die meeste eietydse omskrywings van hierdie veld in. Davies en Elder (2004:1) dui byvoorbeeld ook aan dat “[a]ppplied linguistics is often said to be concerned with solving or at least ameliorating social problems involving language”. Tipiese sosiale vraagstukke wat met taal verband hou, kan

byvoorbeeld wees hoe tale beter onderrig kan word, hoe die opleiding van vertalers en tolke verbeter kan word, hoe geldige taaleksamens of -toetse afgelê kan word, en so meer. Op soortgelyke wyse definieer Schmitt en Celce-Murcia (2010:1) die toegepaste linguistiek as “using what we know about (a) language, (b) how it is learned and (c) how it is used, in order to achieve some purpose or solve some problem in the real world”. In dieselfde trant skryf Hall, Smith en Wicaksono (2011:15-16) dat die toegepaste linguistiek sigself bemoei met “the role language and languages play in perceived problems of communication, social identity, education, health, economics, politics and justice, and in the development of ways to remediate or resolve these problems”.

Hoewel ’n kort historiese oorsig van die verskillende tradisies of paradigmas binne die toegepaste linguistiek in die volgende afdeling aangebied word, is dit belangrik om vooraf daarop te wys dat die huidige opvatting omtrent die aard en omvang van die toegepaste linguistiek as studieterrein aansienlik verskil van die oorspronklike, modernistiese interpretasie daarvan. In ’n poging om met wetenskaplik geloofwaardige oplossings vir taalprobleme vorendag te kom, is die toegepaste linguistiek aanvanklik sonder meer as ’n verlengstuk van die linguistiek beskou. Weideman (2007a:593) verduidelik soos volg:

In the modernist perspective, science is not only the surest knowledge that we have, but the only guarantee of an authoritative solution to a problem. In order to maintain the ‘scientific’ authority of applied linguistics, it is characterised as being at one end of a continuum of theoretical endeavour.

Die beginjare van die toegepaste linguistiek is daarom deur die onkritiese oordrag van taalteoretiese insigte op veral die toegepaste terrein van taalonderrig gekenmerk. Amerikaanse strukturaliste, soos Bloomfield en Lado, se benadering tot die studie van taal was streng beskrywend, aangesien hulle die linguistiek as ’n wetenskap wou ontwikkel. Vir hulle was die primêre taak van die linguïes om taaldata van moedertaalsprekers te versamel en dit dan te analiseer deur ’n studie van die fonologiese en sintaktiese patrone te maak. In navolging van hierdie beskouing, sowel as die destyds heersende behavioristiese leerteorie van Skinner, het die oudiolinguale metode van taalonderrig ontstaan. Op hierdie

manier is taalonderrig nie net met linguistiese analise nie, maar ook met die herhalende inoefening en inkrementele bemeestering van strukturele taaleenhede vereenselwig. In die woorde van Rajagopalan (2004:408) “in its early stages applied linguistics flourished in the shadow of theoretical or general linguistics”. Om hierdie rede is een van die kenmerke van modernistiese definisies van die toegepaste linguistiek dat dit sonder kritiek op ’n kontinuasie tussen die linguistiek en die toegepaste linguistiek aanspraak maak (Weideman, 2007a:593). In die 1980’s definieer Crystal (1987:412) byvoorbeeld die toegepaste linguistiek as “the application of linguistic theories, methods, and findings to the elucidation of language problems that have arisen in other domains”.

In kontras hiermee merk Wei (2011:8) tereg op dat die linguistiek ’n baie klein rol in die hedendaagse konseptualisering van die toegepaste linguistiek speel. Weideman (2007a:594) voer die emansipasie van die toegepaste linguistiek as ’n dissipline in eie reg terug na die ontstaan van die kommunikatiewe benadering tot taalonderrig (“communicative language teaching”). Hierdie ontwikkeling is volgens hom van historiese belang omdat dit die voortsetting van die toegepaste linguistiek as ’n blote onderafdeling van die linguistiek verbreek het.

In the birth of communicative language teaching we find one of the clearest illustrations of the fact that, in designing solutions to language teaching problems, theory does not lead the way. Communicative language teaching was only belatedly justified in terms of second language acquisition research and constructivism. (Weideman, 2007a:594)

Wei (2011:8) wys immers daarop dat huidige kursusse in die toegepaste linguistiek skaars enige verwysing na die linguistiekteorieë van die afgelope twee dekades bevat. Hy skryf trouens dat, in die kontemporêre, postmoderne tradisie van hierdie veld, toegepaste linguiste “feel free to draw on almost any field of human knowledge, from psychological models [...], mathematical models [...] and early Soviet theories of child development such as Vygotsky, to French thinkers such as Foucault and Bourdieu, feminist philosophy, and post-modernism”. Hierdie opmerking van Wei strook grootliks met Weideman (2007a:595) se waarneming dat verskeie postmoderne definisies van die



toegepaste linguistiek die siening vooropstel dat hierdie studiegebied 'n interdisciplinêre aktiwiteit is wat 'n verskeidenheid perspektiewe in die oplossing van taalprobleme kan akkommodeer. In teenstelling met die eng strukturalistiese styl van moderne toegepaste linguistiek word daar dus in die postmodernistiese benadering nie op 'n eenogige beskouing van akademiese navorsing en strewende staatgemaak nie (Weideman, 2013:4483). Gegewe die wantroue in teorieë en ideologieë waardeur dit gekenmerk word “postmodernist perspectives emphasise discontinuity, disaggregation, specificity, fragmentation, and a multiplicity of perspectives” (Weideman, 2007a:595).

Buiten die voorgenoemde, maar in aansluiting daarby, bestaan daar ook 'n tweede vertakking postmoderne definisies wat die sosiale en/of politieke dimensie van die toegepaste linguistiek beklemtoon, en sodoende verantwoordbaarheid ('accountability') as die deurslaggewende kenmerk van toegepaste linguistiese ontwerpe onderskei (Weideman, 2007a; Weideman, 2013). Wei (2011:10) beskryf hierdie uitvloeisel van die postmodernistiese sienswyse as 'n poging om die toegepaste linguistiek as deel van die sosiale wetenskappe te herkonseptualiseer. Vanuit hierdie perspektief word taalkwessies soos taalonderrig, -beplanning en -toetsing as sosio-politiese praktyke bestempel wat potensieel korrup en uitbuitend kan wees (Wei, 2011:10, Weideman, 2013:4483). 'n Belangrike gevolg van hierdie sosiale keerpunt is 'n groeiende bewustheid onder toegepaste linguiste dat versuim om 'n bepaalde standaard in byvoorbeeld tweede- of vreemdetaalonderrig of taaltoetsing te bereik, nie bloot net 'n linguistiese aangeleentheid is nie, maar ook 'n sosio-politiese een (Wei, 2011:10).

Van sy modernistiese begin tot sy postmodernistiese hede het die toegepaste linguistiek dus, soos wat Wei (2011:8) dit stel, merkbare vordering gemaak, van vroeëre pogings om taaloplossings in die strukturele linguistiek te vind tot breër sosiale kwessies wat met taal verband hou, asook 'n kritiese ingesteldheid teenoor hierdie kwessies. Weideman (2007a:597) is nietemin van mening dat beide die

modernistiese en postmodernistiese opvattinge omtrent die toegepaste linguistiek die dissipline verryk. Terwyl modernistiese definisies op die teoretiese, wetenskaplike onderbou van dié veld klem lê, bring postmodernistiese omskrywings (sosiale en politiese) verantwoordbaarheid op die voorgrond. Later in hierdie hoofstuk, wanneer die teoretiese raamwerk bespreek word wat die ontwerp van die onderhawige Afrikaanse nagraadse toets van akademiese geletterdheid ten grondslag lê, word aangetoon hoedat die bydraes van beide hierdie strominge tegelyk in die aanpak van taalverwante probleme erken en gehandhaaf kan word – eersgenoemde (d.i. die modernistiese denkwys) oorwegend in die vorm van konstitutiewe ontwerpvoorwaardes en laasgenoemde (d.i. die postmodernistiese siening) meesal in die gedaante van regulatiewe ontwerpvoorwaardes. Om dit reg te kry, is dit nodig om aan die toegepaste linguistiek as ’n dissipline van ontwerp te dink (Weideman, 2007a) – ’n idee wat later in hierdie hoofstuk in besonderhede toegelig word. Kortom kom laasgenoemde daarop neer dat toegepaste linguistiek hulle nie teen die oënskynlik mededingende definisies van hierdie terrein moet blind staar nie, maar ’n perspektief moet behou wat tot die gemeenskaplike kern van die dissipline deurdring. Gerugsteun deur ’n analise van verskeie modernistiese en postmodernistiese beskrywings van die veld, argumenteer Weideman (2007a) dat die ontwerpelement onderliggend en gemeenskaplik aan beide hierdie interpretasies is. Hy skryf dat “[f]or both modernist and postmodernist interpretations of the field, [applied linguistics] is an endeavour that allows professionals to conceive, plan, and develop designs that address language-related problems” (Weideman, 2013:4480). Hy is trouens van mening dat die “moment of design” die wesenskenmerk van die toegepaste linguistiek is (Weideman, 2007a:589).

Uit die bogenoemde behoort dit duidelik te wees dat om die verwysingsveld van die toegepaste linguistiek af te baken nie ’n eenvoudige saak is nie, aangesien hierdie dissipline in sy ontwikkelingsgang, vanaf sy ontstaan in die 1940’s ná die Tweede Wêreldoorlog, verskillende gestaltes aangeneem het. Hoewel daar

in die voorgaande afdeling reeds aangetoon is hoedat die modernisme en postmodernisme twee tradisionele uiterstes binne die toegepaste linguistiek verteenwoordig, kan die geskiedkundige verloop van hierdie dissipline weliswaar oor sewe tradisies heen gevolg word. 'n Kort oorsig van die verskillende paradigmas binne die toegepaste linguistiek word vervolgens aangebied. In gevalle waar, soos in die hierdie studie, 'n verantwoordelike kultivering van die dissipline aan die orde is, is só 'n historiese terugblik van intrinsieke belang:

[W]here new entrants into the discipline remain unaware of what has preceded their work, they may either uncritically accept current (usually postmodernist) definitions of the field, or, equally uncritically, fall victim to some of the ideological baggage that has historically come with the use of the term 'applied linguistics'. Both situations are undesirable, and restrict rather than open up and liberate any attempt at responsibly developing the discipline of applied linguistics. (Weideman, 2007a:589)

## 2.2 'n Oorsig van die verskillende paradigmas binne die toegepaste linguistiek

Die sistematiese insig dat die toegepaste taalkunde as 'n dissipline gekenmerk deur ontwerp as kernmoment beskryf kan word, moet deur 'n historiese blik op sy ontwikkelingsgeskiedenis aangevul word. Kyk ons na hierdie ontwikkelingsgang, word dit duidelik dat die toegepaste linguistiek in terme van 'n reeks opeenvolgende paradigmas of tradisies beskryf kan word (Weideman 2009b; 2013). Die tabel hieronder dien as 'n verwysingspunt vir die bespreking wat volg (kyk tabel 2.1).

<b>Paradigma of tradisie</b>	<b>Gekenmerk deur</b>
1. Linguisties of behavioristies	“Wetenskaplike” benadering
2. Linguisties uitgebreide paradigma	Taal as ’n sosiale verskynsel
3. Multidissiplinêre model	Gee nie slegs aan taal aandag nie, maar ook aan leerteorie en pedagogie
4. Navorsing in tweedetaalverwerwing	Eksperimentele navorsing oor hoe tale aangeleer word
5. Konstruktivisme	Kennis van ’n nuwe taal word interaktief gekonstrueer
6. Postmodernisme	Politiese oorwegings in onderrig; ’n veelheid van perspektiewe
7. Komplekse sisteem-benadering	Die verskyning van taal as organies en nieliniêr deur dinamiese aanpassing

*Tabel 2.1 Sewe opeenvolgende tradisies binne die toegepaste linguistiek<sup>7</sup>*

Soos vantevore aangedui is, word die eerste tradisie van die toegepaste linguistiek hoofsaaklik deur modernistiese insigte gekenmerk. Ooreenkomstig hierdie denkwysse, waarvolgens ’n sterk teoretiese onderbou en wetenskaplike analise van deurslaggewende belang is, word die toegepaste linguistiek bloot as ’n subterrein van die linguistiek bestempel. Taalonderrig<sup>8</sup> geskied gevolglik aan die hand van strukturalisties geïnspireerde benaderings soos die oudiolinguale metode (wat ook behavioristies van aard is). Verder put dit inspirasie uit die kontrastiewe analise van taalstruktuur ten einde potensiële struikelblokke in die onderrig- en leerproses te identifiseer. Oor die opvoedkundige uitwerking van hierdie tradisie skryf Roets (1999:32) dat “die grootste waarde van die oudiolinguale benadering die oënskynlik wetenskaplike geloofwaardigheid, die gerief en die objektiwiteit by die voorsiening van die nodige gereedskap aan die taalleerder [was], sonder dat hy enige idee gehad het van hoe hy dit moes gebruik”.

In die tweede tradisie word taal wel as ’n sosiale verskynsel erken, maar hierdie tradisie behels in wese, volgens Weideman (1999:81-82), ’n uitbreiding van die

7. Hierdie tabel is ’n vertaalde weergawe van die oorspronklike soos verkry uit Weideman (2013:4481).

8. Hoewel hierdie studie op taaltoetsing (in die vorm van akademiese geletterdheidstoetsing), en nie soseer taalonderrig nie, toegespits is, word daar in die bespreking geredelik na ontwikkelinge op laasgenoemde gebied verwys, aangesien die toegepaste linguistiek sy invloed die sterkste deur hierdie subdissipline laat geld het. In ’n later stadium word aangedui hoedat die ontwikkelingsgeskiedenis van spesifiek taaltoetsing deur die verskillende strominge binne die toegepaste linguistiek (en in samehang daarmee ook deur taalonderrig) beïnvloed is.

voorgenoemde linguistiese paradigma. Die omvang van linguistiese analise wat aanvanklik slegs fonologies, morfologies en sintakties van aard was – ’n soort ontleding wat vir ’n geruime tyd baie prominent in linguistiese werk was – is in hierdie tradisie verbreed om vorme van taalstudie in te sluit wat taal in ’n sosiale konteks plaas, soos semantiekstudies, tekslinguistiek en diskoersanalise. Die Amerikaanse linguis Dell Hymes bevorder byvoorbeeld die idee van “communicative competence” as ’n teenhanger vir Noam Chomsky se “linguistic competence” (Rajagopalan, 2004:404), terwyl Michael Halliday (in Rajagopalan, 2004:405) aankondig dat “[p]robably the most significant feature of linguistics in the seventies is that man has come back into the center of the picture”. Die siening dat taalonderrig nie slegs tot die onderrig van grammatikale strukture en die internalisering van reëls beperk kan word nie, begin gevolglik al hoe meer veld wen (Rajagopalan, 2004:405). Die invloed van hierdie denke op die toegepaste linguistiek manifesteer in die vorm van die ontwikkeling van verskeie kommunikatiewe benaderings tot taalonderrig in die laat sewentigs en vroeë tagtigs (Weideman, 1999:82).

Weideman (1999:82) wys egter daarop dat hierdie ontwikkelinge steeds aan die “eerste generasie”-opvatting dat taal eers geanaliseer moet word alvorens dit onderrig kan word, getrou gebly het, met die gevolg dat “[w]hat all these developments yielded in the end, however, remained a linguistic conception of applied linguistics”. Boonop, hoewel taalfunksies (en nie meer net taalvorme nie) voortaan ’n belangrike rol in taalonderrig sou speel, was dit in hierdie stadium, by die vroeë konsipiëring van die kommunikatiewe benadering tot taalonderrig, nog nie duidelik hóé taalfunksies onderrig en aangeleer behóórt te word nie. Weideman (1999:82) merk in hierdie verband op: “[W]hat was missing – at least initially – in the communicative approach, was a theory of language learning.” Hoewel die tweede tradisie dus ’n verruimde perspektief omtrent taal tot gevolg gehad het, het ’n geskikte leerteorie steeds ontbreek – ’n behoefte wat tot die derde tradisie van die toegepaste linguistiek aanleiding gegee het.

Toegepaste linguïste het begin besef dat daar nie alleenlik op insigte uit die linguïstiek staatgemaak kan word om werkbare oplossings vir taalprobleme te vind nie. Die derde tradisie behels daarom 'n multidissiplinêre model. Rajagopalan (2004:410) verduidelik: “[R]esearchers recognized the need to turn to disciplines such as sociology, anthropology, education, cognitive science, and so forth in addition to linguistics in order to formulate their own theoretical frameworks suited to their applied goals.” Veral sielkunde en opvoedkunde word bruikbare bronne van kennis omtrent kwessies soos hoe tale aangeleer word en watter onderrigstyle doeltreffend werk (Weideman, 2013:4482). Die linguïstiek verloor gevolglik sy sterk houvas op die toegepaste linguïstiek, aangesien oplossings vir taalprobleme voortaan teoreties-interdissiplinêr geregverdig kon word. Hoewel dus steeds verkleef aan die modernistiese sienswyse, begin die interdissiplinêre tradisie om 'n bewussyn van veelvoudige perspektiewe (wat later tekenend van die postmodernistiese perspektief is) wakker te maak (Weideman, 2013:4482).

Die interdissiplinêre styl kom egter nie in die 1980's tot volle reg nie, aangesien navorsing oor tweede taalverwerwing (“second language acquisition”) die botoon begin voer. Dit lui 'n vierde tradisie van toegepaste linguïstiese werk in wat aanvanklik deur Chomsky se postulering van 'n ingebore taalverwerwingsmeganisme, en later deur die kommunikatiewe benadering tot taalonderrig, aangedryf is. Weideman (2013:4482) skryf dat in hierdie tyd “the field saw increasingly sophisticated work on how languages are learned – and importantly, how they are learned in the classroom”. Saam met hierdie tradisie, en in aansluiting daarby, neem veral die konstruktivistiese leerteorie 'n belangrike plek in taalonderrig in en ontwikkel dit tot 'n vyfde paradigma binne die toegepaste linguïstiek – 'n beskouing waarvolgens kennis van 'n nuwe taal interaktief deur middel van kommunikasie met ander gekonstrueer word. Sodoende word die kommunikatiewe aanpak in taalonderrig dus eindelijk van 'n teoretiese onderbou in die vorm van 'n paslike leerteorie voorsien. Hieroor maak Weideman (2013:4482) die volgende opmerking:

In constructivism, incidentally, one found a belated psychological justification for communicative teaching. All of the basic techniques of the communicative approach, viz. information gap exercises, role play tasks and group information gathering techniques, were ideal techniques for allowing the learner to build a language in interaction with others.

Teen die einde van die twintigste eeu verskyn die sesde tradisie wat die postmodernistiese wendingspunt inhou. Soos reeds aangetoon is, beklemtoon hierdie paradigma eerstens dat daar verskeie invalshoeke is vanuit waar taalprobleme benader kan word – ’n uitgangspunt waarvoor die weg reeds in die derde, multidissiplinêre tradisie gebaan is. Verder vestig hierdie paradigma ook die aandag op die sosio-politiese en etiese aspekte van toegepaste linguïstiese werk in ’n poging om die beoefenaars van hierdie dissipline tot groter verantwoordings te roep in die uitvoering van hulle taak (Weideman, 2013:4483).

Hoewel ’n opkomende, sewende tradisie in die gedaante van ’n komplekse sisteem-benadering op die horison blyk te wees, is dit tans nog nie duidelik hoe dit met vorige werkwyses binne die toegepaste linguïstiek saamhang, of in watter nuwe rigting die aanneming daarvan die dissipline sal stuur nie. In hierdie benadering word taal beskou as iets wat geleidelik, maar nie op ’n lineêre of voorspelbare wyse nie, te voorskyn tree (Van Dyk, 2010:39-40; Weideman, 2013:4483). Vir ’n meer volledige bespreking van die kenmerke van dié wordende paradigma raadpleeg Weideman (2009b).

### 2.3 Ontwerp as die wesenskenmerk van die toegepaste linguïstiek

Daar is reeds in die voorafgaande gedeelte uitgewys hoedat die toegepaste linguïstiek in wese met die vind van oplossings vir taalprobleme gemoeid is. ’n Oplossing word tipies in die vorm van ’n ontwerp of plan, soos byvoorbeeld ’n taalkursus, taaltoets of taalbeleid aangebied (Weideman, 2017:213). Met verwysing na die werk van Schuurman<sup>9</sup> voer Weideman (2017:220) aan dat die

---

9. Volgens Weideman (2007b) spruit sy filosofiese raamwerk vir die toegepaste linguïstiek hoofsaaklik voort uit die werk van Egbert Schuurman. Schuurman en sy promotor, Van Riessen, het Dooyeweerd

begrip ‘ontwerp’ as “giving shape to, planning, or forming” verstaan kan word – d.i. ’n tegniese aktiwiteit waaruit ’n tegnies-gestempelde, toegepaste linguistiese artefak of ontwerp te voorskyn kom. Volgens Weideman (2017:220) het tegniese ontwerpe twee terminale of definiërende funksies, naamlik ’n funderings- of basisfunksie en ’n kwalifiserende of leidende funksie, wat die eindpunte beskryf van waar die ontwerpe onderskeidelik begin (of op bou) en hulle aard vind, of eindig. Ooreenkomstig die ontwikkelingsgeskiedenis van die toegepaste linguistiek vind ’n tegnies-gestempelde, toegepaste linguistiese ontwerp in die reël sy fundering in ’n teoreties-wetenskaplike basis of onderbou, en omdat dit in die wetenskap om verdiepte en sistematiese *denke* gaan, kom hierdie funderingsfunksie derhalwe in die analitiese modaliteit van die werklikheid tot uiting. Weideman (2017:220) bied die volgende verklaring vir die analitiese funderingsfunksie van toegepaste linguistiese ontwerpe:

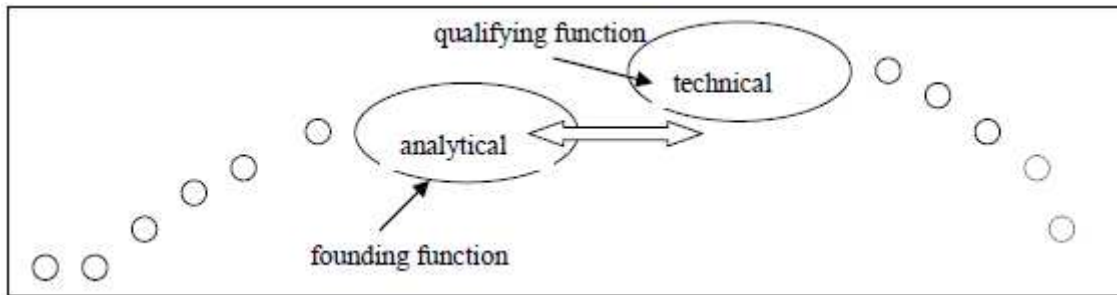
[A] course, test or policy cannot be designed if it is impossible to find some theoretical insight on which to base the design, enabling it to be defended in terms of an analytical perspective, or justified by theory. One might make all manner of plans to solve language problems, but in order to qualify as an applied linguistic plan, it must have made the detour into building a scientific foundation for the [...] design of the artefact.

Omdat ’n toegepaste linguistiese artefak, soos hierbo vermeld, as ’n mensgemaakte, bedinkte plan of oplossing (m.a.w. ’n ontwerp) beskou kan word, is die kwalifiserende of leidende funksie daarvan – wat die *aard* van ’n individuele saak aandui – logieserwys in die tegniese aspek van die werklikheid te vinde – ’n modaliteit wat deur die tegniese vormingsfunksie van die mens getipeer word. Weideman (2017:225) wys trouens met nadruk daarop dat alle toegepaste linguistiese ontwerpe hulle oorsprong in die leidinggewende, tegniese verbeeldingskrag van die toegepaste linguïst (eerder as in die analitiese funderingsfunksie van die artefakte) het – en die toegepaste linguistiek om hierdie rede by uitstek ’n dissipline van ontwerp is. Figuur 2.1 is ’n skematiese voorstelling van hierdie raamwerk wat oorgeneem is uit Weideman (2017:220):

---

en Vollenhoven se filosofiese visie van die modale struktuur van die kosmos (wat binne die *Wysbegeerte van die Wetsidee* ontwikkel is) vir die tegniese wetenskappe hersien en uitgebou.





*Figuur 2.1 Terminale funksies van 'n toegepaste linguistiese ontwerp*

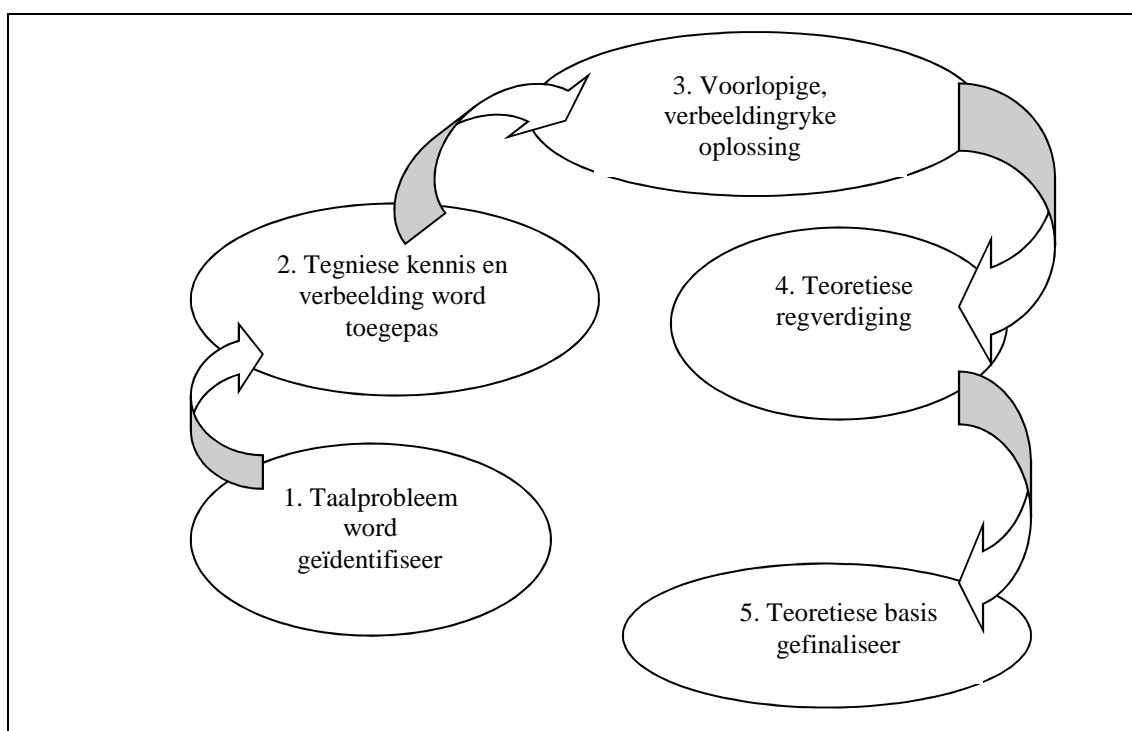
Volgens hierdie interpretasie van die toegepaste linguistiek skryf die teoretiese basis dus nie die ontwerp voor en bepaal of lei dit nie, maar bied eerder 'n (daaropvolgende) rasionaal of regverdiging daarvoor – wat in die geval van 'n taaltoets spesifiek die vorm van 'n toetskonstruk of teoreties-gedefinieerde taalvermoë sou aanneem (Weideman, 2006:72). Sodra die analitiese funderingsfunksie egter tot die tegniese kwalifiseringsfunksie herlei sou word en die teoretiese basis van 'n ontwerp sodoende ten koste van sy tegniese wesensaard beklemtoon word (en andersom), ontstaan paradigmatiese vooroordele wat die ewewig versteur en verantwoordelike ontwerp verydel. Weideman (2017:226) verduidelik soos volg:

Where that [theoretical] justification is given too much prominence, as in the rational defence of audiolinguism, it is indicative of a modernist bias, and all of the contention that that design prejudice carries with it. And where the justification delivers a blind spot in the actual designs of applied linguistic artefacts, as in the case of the postmodernist contestation of modernist directions in our field, there is a comparable deficiency. Responsible design needs to overcome both biases, and it can do so through imaginatively conceived solutions that can be theoretically defended and refined.

Die diagram hieronder (oorgeneem en vertaal uit Weideman, 2009a:244) wys hoedat in die praktyk van toegepaste linguistiese ontwerp dit weliswaar die tegniese verbeelding en kennis van die ontwerper is wat die ontwerpproses rig en kwalifiseer, terwyl die teoretiese regverdiging dikwels eers agterna of terugwerkend verskaf word. Tussen die leidende tegniese funksie en die analitiese funderingsfunksie kom 'n wederkerige of onderlinge verhouding gevolglik tot stand.

That is, in the design of an applied linguistic instrument, the technical imagination of the designer leads the whole endeavour, but at some point in the design process the development of the artefact must open itself up to critical modification and even

correction by analytical and theoretical considerations and rational argument. (Weideman, 2009a:244)



*Figuur 2.2 Vyffases van toegepaste linguistiese ontwerpe*

Die opvatting dat 'n toegepaste linguistiese artefak 'n teoretiese rasonaal moet hê, is volgens Weideman (2017:220) 'n analogiese, tegniese konsep of idee wat ontstaan sodra die leidende tegniese funksie van 'n ontwerp met die analitiese dimensie van die werklikheid in verband gebring word. Met *analogiese konsep* word bedoel 'n teoretiese konseptualisering van die verhouding tussen 'n unieke modaliteit van die werklikheid – in hierdie geval die tegniese aspek – en 'n ander unieke dimensie, soos die analitiese (Weideman, 2009a:243). Laasgenoemde is egter nie die enigste modale aspek waarmee die leidende tegniese funksie van 'n toegepaste linguistiese ontwerp in wisselwerking tree nie. Later in hierdie hoofstuk word aangedui op watter manier die tegniese funksie ook met al die ander dimensies van die werklikheid (bv. die numeriese, ruimtelike, biotiese, en so meer) skakel. Hierdie verbindings is belangrik vanweë die reeds vermelde analogiese konsepte wat daaruit voortvloei, aangesien in elk van hierdie konsepte 'n normatiewe ontwerpbeginsel opgesluit lê wat as voorwaardes vir verantwoordelike (toets)ontwerp geld (Weideman, 2017:224). Saam maak

hierdie beginsels 'n teoretiese raamwerk van konstitutiewe en regulatiewe ontwerpvoorwaardes uit wat weldra volledig in behandeling kom.

Dit is in hierdie stadium eers nodig om die fokus van die bespreking van hierdie meer algemene stellings oor die aard van die toegepaste taalkunde na taaltoetsing as 'n spesifieke subdissipline van die toegepaste linguistiek te verskuif. Daar volg voorts 'n kort historiese opgaaf van die belangrikste benaderings in die ontwikkelingsgang van taaltoetsing, wat, soos duidelik sal blyk, nou saamhang met die onderskeie strominge binne die toegepaste linguistiek wat reeds hierbo bespreek is. Verder word besondere aandag aan die kwessie van geldigheid in toetsontwerp gegee wat as een van die mees fundamentele konsepte in taaltoetsing beskou word en met verloop van tyd op verskeie maniere deur taaltoetsteoretici geïnterpreteer is (Van der Walt & Steyn, 2007:138; McNamara, 2011:353; O'Sullivan, 2011:260). Die teoretiese raamwerk wat vir hierdie studie gebruik word, gee juis 'n eie, unieke vertolking aan geldigheid in die vorm van 'n reeks tegniese ontwerpbeginsele. Om Weideman (2017) se konsepsie van geldigheid te kan verstaan en dit daarna betekenisvol te kan aanwend, is agtergrond omtrent die tradisionele en hedendaagse konseptualisering van hierdie begrip 'n noodsaaklikheid, en daarom word sodanige oorsig verskaf alvorens die raamwerk in sy geheel aangebied word.

### **3. Taaltoetsing: 'n Subdissipline van die toegepaste linguistiek**

#### **3.1 Die ontwikkelingsgang van taaltoetsing**

Sedert die tweede helfte van die twintigste eeu toe taaltoetsing 'n geïnstitutionaliseerde, 'wetenskaplike' praktyk geword het (McNamara & Roever, 2006:1), het verskeie benaderingswyses tot of skole van taaltoetsing verskyn wat veral deur ontwikkelinge op die gebied van taalonderrig beïnvloed is (Roets, 1999:31; Van Dyk, 2010:60; Green, 2014:173).

Die eerste benaderingswyse wat sy bloeitydperk in die 1960's beleef het, was die psigometries-strukturalistiese periode van taaltoetsing (McNamara, 2000:14) wat McNamara en Roever (2006:1) as “an interdisciplinary endeavour between Applied Linguistics and psychometrics” beskryf. Die idee agter die totstandkoming daarvan was om 'n nuwe soort objektiewe assessering daar te stel wat die tradisionele, voorwetenskaplike toetspraktyke van die “grammar translation” metode van taalonderrig kon vervang (Green, 2014:178). McNamara (2000:13-14) verduidelik hoe in aansluiting by die strukturalistiese linguistiek, wat taalkennis aan die bemeestering van die kenmerke van 'n taalsisteem gelykstel, toetse in hierdie tyd toegespits is op kandidate se kennis van die grammatikale sisteem, woordeskat en uitspraak. Samevallend met die behavioristiese, oudiolinguale taalonderrigmetode wat deur “n opeenstapeling van atomiese [taal]deeltjies” (Roets, 1999:32) gekenmerk is, is kandidate se kennis op 'n geatomiseerde en gedekontekstualiseerde manier getoets, en verskillende komponente van taalkennis (soos bv. grammatika en woordeskat) boonop afsonderlik van mekaar, as so te sê selfstandige groothede.

Hierdie toets van aparte, individuele komponente van kennis, ook bekend as diskrete itemtoetsing, is deur teorie en praktyk binne die psigometrie versterk, die destyds opkomende wetenskap van die meet van kognitiewe vermoëns. Omdat psigometriese toetsing veral die lewering van betroubare resultate beklemtoon het, het taaltoetse “hoofsaaklik [...] uit stringe atomiese onderdeeltjies wat ondubbelsinnig reg of verkeerd gemerk kon word” (Roets, 1999:32) bestaan. Objektiewe punttoekenning was belangrik en daarom is veelvuldigekeusevrae bo opstelle wat subjektiewe oordeel vereis, verkies (Green, 2014:178). Van Dyk (2010:62) dui aan dat uit die samewerking van gedragsielkundiges met linguïste “presiese en sogenaamde objektiewe taaltoetse met redelik aanvaarbare statistiese attribute” voortgebring is. Terwyl argumente vir die geldigheid van tradisionele toetse op die kwalifikasies en status van die ontwerpers daarvan gebaseer is, het die psigometriese benadering tot geldigheid die nodigheid van ‘harde’, statistiese bewyse beklemtoon (Green, 2014:178). Die

klem op metingsresultate wat in getalle uitgedruk en verder analiseer kon word, is duidelik.

Hoewel diskrete itemtoetsing dus 'n poging was om (in ooreenstemming met die modernisme) streng wetenskaplik te werk te gaan, het hierdie toetstradisie in die woorde van Van Dyk (2010:62) op “niks meer as 'n beperkte blik op taal” berus nie. Oller (1979:212) merk in hierdie verband op dat “[a]ny discourse in any natural language is more than the mere sum of its analysable parts”. Die ontoereikendheid van die psigolinguistiese benadering in daardie tyd is juis daarin geleë dat dit té uitsluitlik gefokus het op “knowledge of the formal linguistic system for its own sake rather than on the way such knowledge is used to achieve communication” (McNamara, 2000:14-15). 'n Diskrete itemtoets van grammatika kon dalk wel onthul of 'n leerder 'n formele reël kan toepas of nie, maar “could not reveal the way [...] in which that learner might attempt to express an idea in conversation or in a letter” (Green, 2014:185). Ondanks hierdie tekortkominge het die psigometriese fase egter 'n waardevolle en invloedryke bydrae tot die veld van taaltoetsing gelever deur praktyke te help vestig wat tot vandag toe nog 'n fundamentele rol in verantwoordelike toetsontwerp speel (vgl. Weideman se konstitutiewe voorwaardes vir toetsontwerp). McNamara en Roever (2006:2) beaam dat (selfs in die lig van die belang van die sosiale dimensie van taaltoetsing) “adherence to the fundamental rules of what makes a ‘good test’ is indispensable for ensuring the quality of assessments and the defensibility of conclusions drawn from them”.

In reaksie op die gebruik van diskrete itemtoetse en in antwoord op die ontluikende kommunikatiewe beweging in taalonderrig ontstaan daar in die sewentigerjare 'n vraag na 'n meer geïntegreerde benadering tot taaltoetsing (McNamara, 2000:14). Die begin van hierdie behoefte val saam met die linguisties uitgebreide paradigma (kyk tabel 2.1) binne die toegepaste linguistiek en behels 'n fokusverskuiwing “weg van diskrete linguistiese komponente na 'n meer sosiale ingesteldheid waar die uitwerking van linguistiese aktiwiteite op

kommunikasie bepaal kon word” (Van Dyk, 2010:64). Volgens Green (2014:196) betoog Oller gedurende hierdie tyd dat geldige taaltoetse beide geïntegreerd én pragmaties moet wees. Die resultaat hiervan was die ontwikkeling van toetse wat verskillende taalkomponente (uitspraak, grammatika, woordeskat) tesame met mekaar, in plaas van afsonderlik, getoets het, terwyl kandidate tegelykertyd ook van betekenis en konteks bewus moet wees (Green, 2014:196; McNamara, 2000:15). Toetse wat byvoorbeeld die aflê van mondelinge onderhoude, die skryf van tekste, en begrip van uitgebreide gesproke en geskrewe diskoers vereis, is gedurende hierdie periode in gebruik geneem (McNamara, 2000:15).

Verskeie probleme is egter met dié manier van toetsing ondervind. Volgens McNamara (2000:15) het dit spoedig geblyk dat die afneem van geïntegreerde toetse ’n duur en tydrowende praktyk is, dat dit moeilik is om hierdie tipe toetse te bepunt en dat die resultate daarvan potensieel onbetroubaar kan wees, veral in gevalle waar assessore se beoordelings van mekaar verskil. Om moeilikhede soos hierdie die hoof te bied, ondersoek eers Carroll et al. (1959) en later Oller (1979) die gebruik van ’n alternatiewe, meer ekonomies doeltreffende toetsstipe, by name sinsluiting of *cloze testing* – ’n leestoets wat die herstel van ’n teks waarin woorde sistematies weglaat word, behels. Daar is aangevoer dat sinsluiting dieselfde soort vaardighede as sogenaamde “productive tests” toets, aangesien grammatiese, leksikale, kontekstuele en pragmatiese kennis geïntegreer moet word ten einde die ontbrekende woorde te kan verskaf. Sinsluiting was gevolglik ’n baie gewilde vorm van geïntegreerde toetsing in die 1970’s en vroeë tagtigs, en word steeds vandag algemeen gebruik (McNamara, 2000:15). Soos wat sal blyk in die besprekings wat gaan volg oor die ontwerp van TAGNaS en die argumente wat onderliggend is aan ’n goeie gedeelte van daardie ontwerp, kan daar steeds argumenteer word ten gunste van die bruikbaarheid van verbeeldingryke modifikasies van die prosedure van sinsluiting.

Van Dyk (2010:65) beskou myns insiens met reg die soeke na 'n tipe toets wat taal geïntegreerd kon toets as die begin van die oorskakeling na die kommunikatiewe benadering tot taaltoetsing. Hoewel Oller se “unitary competence hypothesis” – naamlik dat verskillende (sub)toetse, soos luisterbegrip, mondelinge onderhoude en grammatika, in wese dieselfde onderliggende vermoë, d.i. taalvaardigheid, meet – nie standgehou het nie, is Green (2014:198) ook van mening dat “the integrative approach [...] did have a lasting effect”. Die nodigheid van 'n meer genuanseerde of verruimde definisie van taalvermoë is gevolglik ingesien (Van Dyk, 2010:65-66). Farhady (1979:349) merk in hierdie verband op dat “not all integrative tests were assessing the communicative abilities of the students, nor were all of them as true to life as had been assumed”. Trouens, al die benaderings tot taaltoetsing tot dusver deel die beskouing dat taal 'n abstrakte sisteem is, verwyderd van die sosiale kontekste waarin dit gebruik word (Green, 2014:198). Met verwysing na Roets (1990), beskryf Van Dyk (2010:67) die klemverskuiwing wat vervolgens in taaltoetsontwerp ingetree het soos volg:

Die fokus het [...] verskuif na die ontwerp van kommunikatiewe taaltoetse wat op 'n gegewe tydstip 'n snit maak om te bepaal wat die huidige stand van 'n persoon se vlak op een of ander taalgebied is ten einde te kan bepaal wat sy / haar potensiaal is om in die werklike lewe taalmatig te oorleef.

McNamara (2000:17) bring hierdie ontwikkeling in taaltoetsing met die verskyning van Hymes (1972) se teorie oor kommunikatiewe bevoegdheid in verband.

The theory of communicative competence represented a profound shift from a psychological perspective on language, which sees language as an internal phenomenon, to a sociological one, focusing on the external, social functions of language.

Volgens hierdie outeur verskil kommunikatiewe van geïntegreerde toetse in dié opsig dat hulle met die sosiale rolle wat kandidate bes moontlik in die werklike lewe sou aanneem (bv. as akademiese gebruiker, d.w.s. as voorgraadse of nagraadse student, of as iemand wat in die sakewêreld moet kommunikeer, of bloot sosiaal), rekening hou, en 'n manier bied om die eise wat hierdie rolle stel in besonderhede aan te gee (McNamara, 2000:17). Voorstanders van

kommunikatiewe toetsing het aangedring op die outentisiteit van hulle toetsmateriaal met die doel om lewensgetroue omstandighede in hulle toetse te reproduseer (Green, 2014:199). Die ontwerp van sodanige toetse het taaltoetsontwerpers egter aanvanklik in 'n worstelstryd gewikkel gehad (Van Dyk, 2010:67), soos Green (2014:199) verder tereg opmerk:

The point was not to test knowledge of language systems, whether as discrete components or integrated whole, but the ability to use language functionally to carry out real world tasks: to understand and pass on ideas, to form and maintain relationships, to negotiate transactions.

Green (2014:202) wys egter vermanend op Bachman se standpunt dat die kommunikatiewe klem wat op die suksesvolle uitvoering van lewenswerklike take geplaas word, nie in die plek kan staan van 'n duidelike definisie van die kennis, vaardighede of vermoëns wat geassesseer word nie. As antwoord op die uitdaging van kommunikatiewe toetsing, asook die voorgenoemde “behoefte aan 'n [...] sosiolinguisties-regverdigbare benadering tot taaltoetsing” (Van Dyk, 2010:67), ontwikkel verskeie teoretici sogenaamde “models of communicative ability” wat poog om die onderskeie komponente van kommunikatiewe bevoegdheid in 'n taal (of tweede taal), nader te spesifiseer. McNamara (2000:17) skryf dat “the challenge of communicative language testing was matched by a continuing theoretical engagement with the idea of communicative competence and its implications for the performance requirement of communicative language testing”.

Kommunikatiewe taaltoetsing is byvoorbeeld aanvanklik ooreenkomstig 'n model van Michael Canale en Merrill Swain wat in die vroeë 1980's ontwikkel is, uitgevoer. In hulle model word kommunikatiewe bevoegdheid omskryf as bestaande uit grammatiese, sosiolinguistiese, strategiese en diskoersbevoegdheid. Hierdie model is egter later van tyd deur navorsers soos Lyle Bachman en Adrian Palmer uitgebrei en geherorganiseer. Bachman en Palmer (1996) se omskrywing van taalvermoë kom spesifiek in Hoofstuk 3 van hierdie studie ter sprake, waar die ontwikkeling van die konstruk van akademiese geletterdheid waarop die betrokke nagraadse akademiese geletterdheidstoets wat



die fokus is van hierdie studie, gebaseer word, onder bespreking kom. Die Bachman en Palmer-model (1996) is hier van belang omdat dit een van die oorspronklike boustene van die konstruksie vir akademiese geletterdheidstoetsing is waarop die ontwerp van die Test of Academic Literacy Levels (TALL), Toets van Akademiese Geletterdheid (TAG) en Test of Academic Literacy for Postgraduate Students (TALPS), asook van die onderhawige Afrikaanse nagraadse toets, gebaseer is. Soos wat later aan die lig sal kom, is hierdie konstruksie nie van die grond af nuut geskep nie, maar dit vloei gedeeltelik voort uit die werk van Nan Yeld en die Alternative Admissions Research Project (AARP) van die Universiteit van Kaapstad wat op hulle beurt weer die oorspronklike Bachman en Palmer-konstruksie vir die Suid-Afrikaanse hoërsonderwyskonteks hervertolk het (Van Dyk & Weideman, 2004a).

Oor die nut van modelle van kommunikatiewe bevoegdheid skryf McNamara (2000:17-18) dat dit met betrekking tot kommunikatiewe taaltoetsing “a comprehensive framework for test development and testing research, and a basis for the interpretation of test performance” daarstel. Hy is trouens van mening dat konstruksie van taalvaardigheid, tesame met die kwessie van geldigheid, twee van die mees fundamentele konsepte in taaltoetsing is (2011:353). Volgens Van Dyk (2010:69) daag hedendaagse taaltoetsing ontwerpers nie net uit om ’n kommunikatiewe siening van taal in hulle toetsontwerpe te inkorporeer nie, maar ook om die items wat in die toets verskyn met die taalsituasie wat hulle teiken, te belyn. In die praktyk blyk laasgenoemde egter geen maklike taak te wees nie, soos wat McNamara (2000:20) tereg opmerk:

[I]t is [...] true that attempts to apply a complex framework for modelling communicative language ability directly in test design have not always proved easy, mainly because of the complexity of the framework. This has sometimes resulted in a rather tokenistic acknowledgement of the framework and then a disregard for it at the stage of practical test design.

In Hoofstuk 3 (en verder) word egter aangedui hoedat daar in hierdie studie nadruklik pogings aangewend is om te verhoed dat ’n oppervlakkige toepassing van kommunikatiewe taaltoetsing in die ontwerp van die betrokke akademiese geletterdheidstoets neerslag vind, en verantwoordelike toetsontwerp sodoende,

in terme van tegniese belyning<sup>10</sup> tussen toetskonstruk, -spesifikasies en -items, ondermyn word. Dit word later duidelik hoedat die ontwerp van die onderhawige toets, en die konstruk van akademiese geletterheid waarop dit gebaseer is, op 'n verruimde, kommunikatiewe siening van taal berus, hoedat die operasionalisering van die konstruk (d.i. die ontwikkeling, verfyning en seleksie van toetsspesifikasies, subtoetse en taaktipes) deurgaans gedagtig aan die vereistes van belyning en produktiwiteit uitgevoer is (vgl. Van Dyk & Weideman, 2004b), en hoe die akademiese taalgebruik wat getoets word, altyd die analities-gestempelde interaksie wat kenmerkend van daardie soort diskoers is, in gedagte hou. Linguale interaksie bied die basis van kommunikasie met ander, ook wanneer dit tipies gestempel word as akademiese kommunikasie (as onderskeibaar van ekonomiese, sosiale, juridiese of ander tipiese taalgebruik).

### 3.2 Geldigheid in taaltoetsing: Die tradisionele versus hedendaagse perspektief

In taaltoetsing is geldigheid, benewens betroubaarheid, een van die vernaamste kriteria waaraan voldoen moet word (Fulcher & Davidson, 2007; Van der Walt & Steyn, 2007; Van der Walt & Steyn, 2008; Chapelle, 2012; McNamara, 2011; Van Dyk, 2013). Trouens, Fulcher en Davidson (2007:3) beskryf geldigheid as “the central concept in [language] testing and assessment”, terwyl Van der Walt en Steyn (2007:138 & 2008:91) reken dat dit een van die belangrikste oorwegings in die ontwikkeling en evaluering van taaltoetse is. Die geskiedenis van die toegepaste linguistiek getuig egter van 'n worsteling met hierdie begrip wat deur die jare heen op verskeie maniere deur teoretici geïnterpreteer, en veral deur die denke van die Amerikaanse sielkundige Samuel Messick (1980, 1981, 1988, 1989), beïnvloed is. Chapelle (2012:22) merk op dat daar reeds sedert die 1960's – toe die eerste boeke oor moderne taaltoetsing verskyn het – oor

---

10. Tegniese belyning of harmonisering is 'n analogiese konsep wat voortvloei uit die wisselwerking tussen die leidende tegniese funksie van 'n toegepaste linguistiese ontwerp en die estetiese dimensie van die werklikheid. Raadpleeg in hierdie verband Weideman (2017) se raamwerk van ontwerpbeginsels wat aan die einde van hierdie hoofstuk volledig uiteengesit word.

geldigheid geskryf is, en verskillende vertolkings sedertdien aan hierdie konsep gegee is. In haar historiese oorsig identifiseer Chapelle (2012:22-26) onder andere drie belangrike konsepsies van geldigheid wat oor die afgelope vyftig jaar die lig gesien het, en, wat ooreenkomstig die klassifikasie van Van Dyk (2013) breedweg beskou in twee paradigmas, te wete 'n tradisionele en hedendaagse ortodoksie, ingedeel kan word.

Ingevolge die eerste, tradisionele siening van geldigheid is 'n toets geldig wanneer dit meet wat dit veronderstel is om te meet. Hiervolgens is geldigheid 'n inherente eienskap waarvoor 'n toets kan beskik (Van der Walt & Steyn, 2007:139, Van Dyk, 2013:159). Chapelle (2012:22) dui aan dat dit in hierdie benadering hoofsaaklik om die beantwoording van twee vrae gaan, naamlik (1) Wat presies meet die toets? en (2) Hoe goed meet die toets? Om antwoorde op dié vrae te verskaf, is drie onderskeie tipes geldigheid aan die orde gestel en as maatstaf in geldigheidsondersoeke gebruik, by name kriterium-gerigte geldigheid (wat samevallende of konkurrente en voorspellingsgeldigheid insluit), inhoudsgeldigheid en konstruktiegeldigheid, waarvan laasgenoemde oorspronklik deur Cronbach en Meehl (1955) in gebruik gebring is. Fulcher en Davidson (2007:4) skryf dat in die beginjare van geldigheidsondersoeke “validity was broken down into three ‘types’ that were typically seen as distinct. Each type of validity was related to the kind of evidence that would count towards demonstrating that a test was valid”. Op dié wyse het Lado byvoorbeeld ten opsigte van kriterium-gerigte geldigheid aangevoer dat 'n taaltoets geldig verklaar kan word indien die punte daarin behaal redelik met dié van 'n ander geldige toets of kriterium korreleer (Xi, 2008:177-178). Verder word betroubaarheid, d.i. interne konsekwentheid, in die tradisionele benadering nie net as afsonderlik van geldigheid beskou nie, maar ook as 'n noodsaaklike voorwaarde daarvoor (Van der Walt & Steyn, 2007:139).

Van Dyk (2013:158-159) wys daarop dat die klem in hierdie paradigma sterk op die empiriese en bewysbare (“dit wat gemeet en bewys kan word”) val. Van der Walt en Steyn (2007:139) verduidelik soos volg:

The traditional approach reflects a positivistic paradigm, which assumes that a psychologically real construct or attribute exists in the minds of the test takers – this implies that if something does not exist, it cannot be measured.

Hierdie benadering is egter mettertyd as “beperkend”, “nie genuanseerd genoeg nie” en “selfs simplisties” deur taaltoetsteoretici bestempel (Van Dyk, 2013:160). Xi (2008:178) som die problematiek rondom die vroeë, tradisionele konsepsie van geldigheid soos volg op:

[E]arlier conceptualizations of validity [...] focused on a few limited types of validity that support primarily score-based predictions, rather than theoretically and empirically grounded explanations of scores that provide the basis for predictions. Treating validity as different types invited researchers to select only one type as sufficient to support a particular test use. Further, test-taking processes and strategies, and test consequences were not examined.

In aansluiting hierby skryf Davies en Elder (2005:799) dat “classical validity studies looked in different and unrelated directions”. Die geldigverklaring van ’n taaltoets kon dus nie net enigeen van verskeie ‘komponente’ van geldigheid daarvoor gebruik nie, maar hoef ook geen samehang te gesoek het tussen die onderskeie komponente van geldigheid nie.

In die 1980’s ontwikkel daar ’n tweede konsepsie van geldigheid wat die tradisionele definisie van drie tipes geldigheid met ’n enkel, verenigde visie van hierdie begrip vervang, en as’t ware “’n teenreaksie op die oordrewe klem van die bewysbare” was (Van Dyk, 2013:159). Volgens hierdie “single unified view of validity” is geldigheid tegelyk ’n unitêre, maar ook ryk geskakeerde konsep met konstruktgeldigheid as kernkomponent, terwyl kriterium-gerigte en inhoudsgeldigheid deel van konstruktgeldigheid uitmaak (Van der Walt & Steyn, 2007:139). Hierdie siening verteenwoordig die hedendaagse ortodoksie met betrekking tot geldigheid in taaltoetsing en spruit veral uit die werk van Messick wat in 1989 ’n hoogs invloedryke hoofstuk oor geldigheid in *Educational Measurement* publiseer. McNamara en Roever (2006:9-12) wys in hierdie

verband ook op die belangrike bydrae van Cronbach en Meehl wat reeds in 1955 aan die konsep van konstrugeldigheid, as 'n alternatief vir kriterium-gerigte geldigheid, gestalte gee.

Construct validity is ordinarily studied when the tester has no definite criterion measure of the quality with which he is concerned and must use indirect measures. Here the trait or quality underlying the test is of central importance, rather than either the test behaviour or the scores on the criteria. (Cronbach & Meehl, 1955:283)

Sedert die bekendstelling van hierdie term het, soos McNamara en Roever (2006:10) dit stel, “the increasingly central role taken by construct validity, which has subsumed other types of validity” tot een van die belangrikste ontwikkelinge in navorsing oor geldigheid gegroei. So ook die mening dat geldigheid nie bloot 'n wiskundige eienskap soos diskriminasie of betroubaarheid is nie, maar eerder 'n geval van oordeel. Soos vervolgens aangetoon word, met spesifieke verwysing na die idees van Messick, vind daar in hierdie paradigma “'n klemverskuiwing na die interpretatiewe kant van toetsresultate plaas” (Van Dyk, 2013:160).

Messick (1989:13) omskryf geldigheid as “an overall evaluative judgment of the degree to which evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores” – 'n hernieude konseptualisering waarin, volgens Chapelle (2012:23-24), die volgende vier eienskappe met betrekking tot geldigheid opgesluit lê:

- Geldigheid is nie 'n kenmerk van 'n toets self nie; dit is eerder die interpretasies en gebruike van toetse wat as minder of meer geldig beoordeel kan word;
- Geldigheid kan ten beste as een enkele konsep bedink word met konstrugeldigheid as kernelement, in plaas van as bestaande uit veelvoudige tipes geldigheid;
- Geldigheid omvat die relevansie en bruikbaarheid, waarde-implikasies en sosiale gevolge van toetsing. Hierdie beskouing is in kontras met die opvatting dat geldigheid slegs na tegniese oorwegings verwys.

- Die komplekse siening van geldigheid beteken dat validasie 'n deurlopende proses van ondersoek is. Dit is in teenstelling met die tradisionele, produk-georiënteerde perspektief van 'n geldig verklaarde toets waarvoor die navorsing voltooi is.

Messick (1989) stel by die verduideliking van sy idees sy veelbesproke skematiese voorstelling van die fasette van geldigheid bekend, wat trapsgewys 'n aanduiding gee van die tipes navorsing wat met geldigheidsondersoeke gepaard behoort te gaan. Hy benader hierdie aspekte as “holding together within a unified theory of validity” (McNamara & Roever, 2006:13).

	TEST INTERPRETATION	TEST USE
EVIDENTIAL BASIS	Construct validity	Construct validity + Relevance/utility
CONSEQUENTIAL BASIS	Value implications	Social consequences

*Tabel 2.2 Messick se fasette van geldigheid*

Met sy skema benadruk Messick in die eerste plek die belang van bewyse ter ondersteuning van die vertolkings wat aan toetsresultate gegee word in terme van die voorgestelde konstruk(te) wat gemeet word. Die noodigheid van toetskonstrukte wat relevant en bruikbaar ten opsigte van die toetskonteks moet wees, word ook beklemtoon (McNamara, 2011:354). In die tweede plek gee Messick eksplisiet aan die sosiale dimensie van toetsing aandag – 'n aspek wat 'n reeks nuwe, relatief onbekende vereistes ten opsigte van toetsgeldigheid aan die orde laat kom het. Messick beweer nadruklik dat waardeoordele onontkombaar deel van alle toetspunt-interpretasies uitmaak, en staan verder die idee voor dat die konsekwensiële geldigheid van 'n toets, d.i. die sosiale impak van 'n toets op die individu en gemeenskap, 'n noodsaaklike oorweging in die geldigmaking daarvan moet wees (McNamara, 2001:354). McNamara en Roever (2006:12) ekspliseer hierdie standpunte van Messick soos volg:

[Messick] introduced the social more explicitly [...] by arguing two things: that our conceptions of what it is that we are measuring and the things we prioritize in measurement, will reflect values, which we can assume will be social and cultural in origin, and that tests have real effects in the educational and social contexts in which

they are used and that these need to be matters of concern for those responsible for the test.

Davies en Elder (2005) bestempel Messick se bydrae as die begin van “the social turn in the conception of validity in language assessment” en Van der Walt en Steyn (2007:139) merk op dat die sosiale gevolge van toetse sedertdien as ’n “central part of validity” erken word. Dit sluit die verrekening in van kwessies soos die impak van toelatings- of plasingstoetse op die lewens van kandidate, die terugspoeleffek van toetse (‘washback’), etiese oorwegings, die toetskonteks en administratiewe prosedures, asook die karaktereenskappe van diegene wat toets skryf (Van der Walt & Steyn, 2007:139-140). Chapelle (2012:24) skets die uitdaging wat Messick hierbo rig as een van ’n kritiese ingesteldheid, d.i. “to include critical forms of inquiry [...] in order to be able to analyze the values and social consequences underlying test interpretation and use”. Hierdie benaderingswyse vorm die kern van die huidige “critical language testing”-beweging in taaltoetsing, met sy postmodernistiese inslag, waarvan teoretici soos Shohamy en Pennycook leidende figure is.

Hoewel steeds invloedryk, bevraagteken sommige teoretici egter die operasionaliseerbaarheid van Messick se raamwerk, en takseer dit as té vaag, ingewikkeld en veeleisend vir die praktyk. Chapelle (2012:24) skryf:

Indeed, for some researchers ... the scope of concerns raised for validation by Messick’s paper was simply too much. With users needing test scores for decision making, how could test developers and researchers be charged with a never-ending agenda encompassing both the scientific questions about construct interpretations and the ethical questions of values and social consequences? From a pragmatic perspective of getting the validation job done, Messick’s framework made validation seem unapproachably complex for some.

In ’n poging om sodanige kompleksiteite uit die weg te ruim en konseptuele klaarheid te bekom in die vorm van “concepts that can be operationalized and procedures that can actually be accomplished” (Chapelle, 2012:24), voer verskeie teoretici Messick se verenigde benadering tot geldigheid verder en herinterpreteer dit met die klem op verskillende aspekte. So is daar byvoorbeeld McNamara en Roever (2006:14) se hervertolking wat spesifiek ten doel het om

op 'n eenvoudiger manier te wys hoe Messick se raamwerk, op basis van die nodige bewyse, ook van die sosiale dimensie van toetsing kennis neem.

	WHAT TEST SCORES ARE ASSUMED TO MEAN	WHEN TESTS ARE ACTUALLY USED
USING EVIDENCE IN SUPPORT OF CLAIMS: TEST FAIRNESS	What reasoning and empirical evidence support the claims we wish to make about candidates based on their test performance?	Are these interpretations meaningful, useful and fair in particular contexts?
THE OVERT SOCIAL CONTEXT OF TESTING	What social and cultural values and assumptions underlie test constructs and hence the sense we make of scores?	What happens in our education systems and the larger social context when we use tests?

*Tabel 2.3 McNamara en Roever (2006) se herinterpretasie van Messick se fasette van geldigheid*

'n Ander bekende herinterpretasie is die een van Bachman en Palmer (1996) wie se idee van bruikbaarheid of “test usefulness”, volgens Chapelle (2012:24-25), 'n derde verneme beskouing van geldigheid in die ontwikkelingsgang van hierdie konsep verteenwoordig. Hierdie outeurs poog om die gaping tussen teorie en die praktyk te oorbrug deur aspekte van geldigheid by 'n model van toetsbruikbaarheid, wat meer konkrete vereistes vir die evaluering van taaltoetse daarstel, in te lyf. Chapelle (2012:25) verduidelik:

Bachman and Palmer (1996) aimed to interpret validity theory in a manner that could be understood and used by those responsible for developing language tests and justifying their use. Framing the issue as an evaluation of usefulness (rather than justifying interpretations and uses of test scores), [they] were able to communicate [...] that tests have to be evaluated in view of the particular uses for which they are intended.

Usefulness =	Reliability + Construct validity + Authenticity + Interactiveness + Impact + Practicality
--------------	---

*Tabel 2.4 Bachman en Palmer (1996) se model van toetsbruikbaarheid*

Die konsepte wat Bachman en Palmer onder die term bruikbaarheid of nuttigheid verenig, sluit betroubaarheid, konstrugeldigheid, outentisiteit, interaktiwiteit, impak en praktiese sin in. Kortliks behels betroubaarheid (‘reliability’) interne metingskonsekwentheid, terwyl konstrugeldigheid (‘construct validity’) verwys na die geskiktheid van die gevolgtrekkings wat oor kandidate se



vermoëns op grond van toetsresultate gemaak word. Outentisiteit ('authenticity') dui op die mate van ooreenstemming tussen die items in 'n toets en die werklike taaktipes buite die toetskonteks. Interaktiwiteit ('interactiveness') verwys na die mate waartoe toetskandidate se taalvermoë en agtergrondskennis by die voltooiing van 'n toets betrek word. Met impak ('impact') word bedoel die positiewe of negatiewe gevolge wat 'n toets vir individue en instellings kan inhou, terwyl praktiese sin ('practicality') met die implementering van 'n toets en die beskikbaarheid van hulpbronne te make het (Chapelle, 2012:25; Fulcher & Davidson, 2007:15).

Teenoor die toereikendheid van herinterpretasies, soos onder andere die bostaande een van Bachman en Palmer (1996), staan Weideman (2012) egter krities. Hy bevraagteken die oënskynlik blywende opvatting dat maatstawwe vir toetsontwerp, in navolging van Messick se "unified theory of validity", onder 'n enkele sambreelwoord, soos geldigheid, tuisgebring behoort te word – ten spyte daarvan dat dit tot konseptuele troebelheid blyk aanleiding te gee.

Why should conditions for responsible test design continue to be subsumed under 'validity'? [...] [W]e achieve no greater conceptual clarity when we conflate the various design conditions that apply to tests. Far from helping us to reinterpret validity in order to clarify it, such reinterpretation may instead confuse. (Weideman, 2012:8)

In dieselfde trant wys Van Dyk (2013:164) byvoorbeeld daarop hoedat Bachman en Palmer "sonder om direkte kritiek op Messick uit te spreek" bloot subtiel die klem verskuif van "n vereniging van onderskeibare konsepte onder die begrip geldigheid" na "n vereniging van konsepte onder die begrip nuttigheid". Hoewel Bachman en Palmer skynbaar, anders as Messick, die saamgevoegde konsepte van gelyke waarde skat, en nie een (bv. konstruktgeldigheid) bo 'n ander verhef nie, is Van Dyk (2013:164) ook van mening dat "konseptuele skerpheid tot 'n mate verlore gaan in 'n verenigde benadering en dat enige verenigende begrip derhalwe tog as superieur tot die ander geag kan word, in hierdie geval die begrip nut". Ook Fulcher en Davidson (2007:15) meen dat Bachman en Palmer se hervertolking nie werklik konseptueel 'n deurbraak maak nie. Hulle voer aan:

The notion of test 'usefulness' provides an alternative way of looking at validity, but it has not been extensively used in the language testing literature. This may be because

downgrading construct validity to a component of ‘usefulness’ has not challenged mainstream thinking since Messick.

Weideman (2012) maak verder ook beswaar daarteen dat geldigheid, volgens die hedendaagse ortodoksie, slegs van die interpretasie van toetsresultate afhanklik is en nie ’n eienskap van ’n toets nie. Die gevaar hieraan verbonde is dat die gehalte van die meetinstrument self onderspeel kan word. Weideman (2012:4) maan dat “[n]o amount of interpretation can improve the measurement result (score) obtained from an inadequate instrument that gives a faulty and untrustworthy reading”. Weideman (2009a:242-243) stem saam dat toetsresultate op sigself, sonder menslike interpretasie, geen betekenis het nie, maar beklemtoon in dieselfde asem dat subjektiewe interpretasies desnieteenstaande op grond van objektiewe metings gemaak moet word. Daar behoort daarom, volgens Weideman, ’n onderskeid tussen die subjektiewe proses van validasie (die opbou van ’n geldigheidsargument) en die objektiewe geldigheid van ’n toets gemaak te word.

Viewed subjectively, validity is the achievement of validation. Viewed objectively, it is a function of test scores. If the latter were not the case, we would not have been able to ascribe or impute an adequate interpretation to such scores – for those scores would lack not only validity, but also interpretability. (Weideman, 2009a:243)

Omdat “[i]nterpreteerbaarheid nie die probleme uit ’n objektiewe meetinstrument [kan] verwyder nie” (Van Dyk, 2013:165), bly die tegniese krag of toereikendheid van ’n taaltoets, d.i. geldigheid in die gedaante van ’n objektiewe, meetbare eienskap, dus vir Weideman (2012:4) ’n fundamentele oorweging in verantwoordelike toetsontwerp. Voorts ondervang Weideman (2009a, 2012, 2017) sy twee besware teen die hedendaagse sienswyse oor geldigheid (soos hierbo uiteengesit) deur ’n alternatiewe herinterpretasie van Messick se formulerings ter tafel te lê. Hierdie herversolking word vervolgens aangebied en, daaruit voortvloeiend, die volledige teoretiese raamwerk van konstitutiewe en regulatiewe ontwerpvoorwaardes wat aan hierdie studie ten grondslag lê.

### 3.3 'n Alternatiewe perspektief: 'n Balans tussen die konstitutiewe en regulatiewe voorwaardes vir toetsontwerp

Vir Weideman (2012:5) is die begrippe toereikendheid ('adequacy') en gepasheid ('appropriateness') nie net sleutel terme in Messick se omskrywing van geldigheid<sup>11</sup> nie, maar ook twee afsonderlike en onderskeibare konsepte.

Appropriateness is [...] a concept [...] that captures the analogical social dimension of the technically qualified design of language tests. Adequacy, on the other hand, is a concept that is linkable directly to the effects (or effectiveness) of applying a technical instrument such as a language test. (Weideman, 2012:5)

Met hierdie konsepte as uitgangspunt hergiet Weideman (2012) Messick se raamwerk met die doel om 'n moontlike alternatief tot die bestaande herinterpretasies daarvan te verskaf (kyk tabel 2.5).

	adequacy of	appropriateness of
inferences made from test scores	depends on multiple sources of empirical evidence	relates to impact considerations / consequences of tests
the design decisions derived from the interpretation of empirical evidence	is reflected in the usefulness / utility or (domain) relevance of the test	will enhance and anticipate the social justification and political defensibility of using the test

*Tabel 2.5 Die verband tussen enkele fundamentele oorwegings in taaltoetsing*

Hierdie skema behoort, volgens Weideman (2009a, 2012) in die vorm van 'n stel vereistes met betrekking tot taaltoetsing gelees te word (kyk tabel 2.6).

(1)	The technical adequacy of inferences made from test scores depends on multiple sources of empirical evidence.
(2)	The appropriateness of inferences made from test scores relates to the detrimental or beneficial impact or consequences that the use of a test will have.
(3)	The adequacy of the design decisions derived from the interpretation of empirical evidence about the test is reflected in the usefulness, utility or relevance to actual language use in the domain being tested.
(4)	The appropriateness of the design decisions derived from the interpretation of empirical evidence about the test will either undermine or enhance the social justification for using the test, and its public or political defensibility.

*Tabel 2.6 Vereistes vir taaltoetsing*

11. Soos reeds vermeld, definieer Messick (1980, 1981) geldigheid as "an overall evaluative judgment of the adequacy and appropriateness of inferences drawn from test scores".

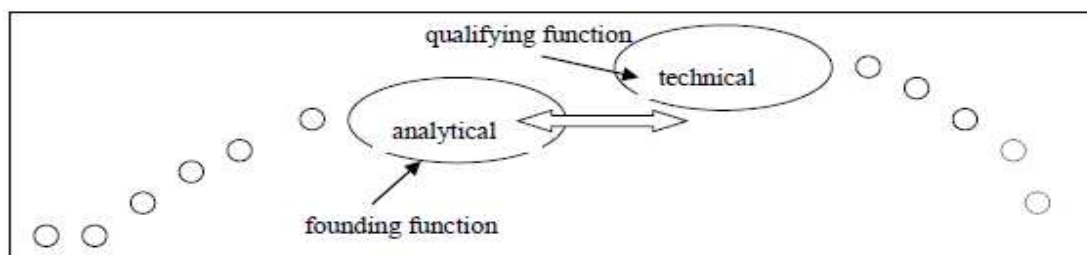
Weideman (2012) beklemtoon dat, hoewel bogenoemde voorstelling op Messick se argument gebaseer is, dit geensins as 'n geldigheidsmatriks, soos in die geval van die oorspronklike, geïnterpreteer moet word nie. Ewemin handel die stellings of uitsprake wat daaruit voortkom slegs oor geldigheid. Vir hom beeld hierdie raamwerk eerder die verband tussen 'n beperkte aantal fundamentele, onderskeibare konsepte in taaltoetsing uit wat nie onder 'n enkel oorheersende begrip, te wete *geldigheid*, verenig behoort te word nie (hoewel sommige van hierdie konsepte steeds met die tegniese krag van 'n toets te make het en daarom aspekte van geldigheid kan insluit):

While obliquely related to the technical power of a test, [the above-mentioned statements] instead articulate the coherence or systematic fit of a number of concepts relating to language testing. What is more, they also articulate some social dimensions of language testing, particularly the social appropriateness, impact, benefits of and public accountability for tests. [...] Surely concepts like technical adequacy, appropriateness, the technical meaningfulness (interpretation) of measurements (test scores), utility, relevance, public defensibility and the like must be conceptually distinguishable to make sense? (Weideman, 2012:7-8)

Weideman (2012) stel derhalwe voor dat daar afstand gedoen moet word van Messick se verenigde visie van geldigheid en dat in die plek daarvan 'n idee van verantwoordelike toetsontwerp verken word wat in 'n raamwerk van konstitutiewe en regulatiewe ontwerpbeginsels tot uitdrukking kom. Sou geldigheid vanuit die perspektief van hierdie raamwerk gedefinieer word, kan dit, soos vervolgens sal blyk, tegelykertyd geag word as 'n objektiewe, meetbare eienskap van 'n toets én 'n konstitutiewe konsep wat deur verdere regulatiewe idees betreffende taaltoetsing ontsluit kan word in die subjektiewe proses van geldigmaking. Laasgenoemde stel idees maak deel uit van hierdie voorstel omdat daar, soos reeds aangetoon is, grondige redes aan te voer is vir die inagneming van sosiale aspekte in geldigheidsondersoeke. Volgens Van Dyk (2010:201) bring Weideman (2017) se raamwerk juis elemente uit beide die tradisionele en hedendaagse paradigmas byeen “wat objektief sal bydra tot geldigheid, maar in balans met mekaar en as deel van die subjektiewe proses van validering”.

Hierdie raamwerk wat die teoretiese uitgangspunt van hierdie studie vorm, word vervolgens bespreek.

Daar is vroeër in afdeling 2.3 gronde aangevoer vir die tipering van die toegepaste linguistiek as 'n dissipline van ontwerp. Daar is spesifiek aangedui hoedat toegepaste linguistiese ontwerpe, soos taaltoetse, oor twee terminale of definiërende funksies, by name 'n leidende tegniese funksie en analitiese funderingsfunksie, beskik (kyk weer figuur 2.1, hieronder verskaf). Soos reeds vermeld, is die gedagte dat 'n toegepaste linguistiese artefak 'n teoretiese rasionaal moet hê, 'n analogiese konsep wat uit die wedersydse inwerking van hierdie twee funksies op mekaar, voortvloei.



*Figuur 2.1 Terminale funksies van 'n toegepaste linguistiese ontwerp*

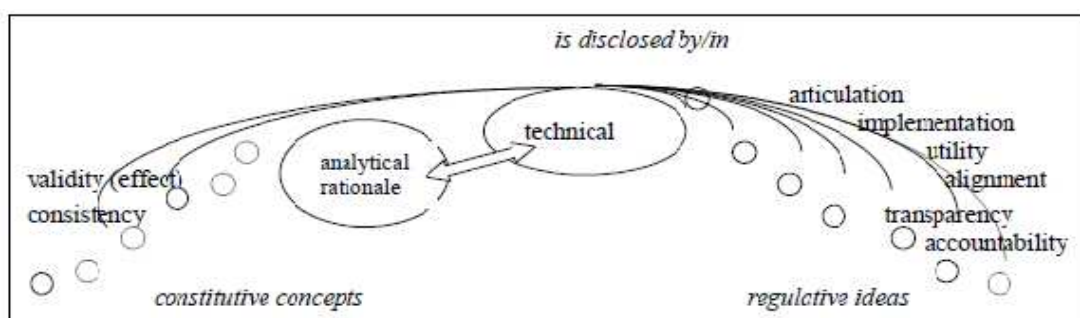
Bo en behalwe die analitiese, skakel die leidende tegniese funksie van 'n toegepaste linguistiese ontwerp egter ook met ander dimensies van die werklikheid. Uit hierdie verbindings spruit voort 'n hele reeks analogiese momente wat elk aan 'n uniek onderskeibare normatiewe ontwerpbeginsel of -voorwaarde uitdrukking gee. Die analogiese momente wat op hierdie manier ontstaan, kan, volgens Weideman (2017:221), in twee kategorieë verdeel word, naamlik 1) konstitutiewe, tegniese konsepte en 2) regulatiewe, linguistiese idees (kyk figuur 2.3 en tabel 2.7).

Konstitutiewe konsepte ontspruit uit die verbinding van die tegniese funksie met die modale aspekte wat dit voorafgaan, en waarop hierdie leidende funksie (m.a.w. die tegniese aspek van ontwerpe) gefundeer is, d.i. die numeriese, ruimtelike, kinematiese, fisiese, organiese, biotiese, psigiese en analitiese dimensies. So byvoorbeeld is betroubaarheid (interne konsekwentheid) 'n konstitutiewe konsep wat spruit uit die wisselwerking tussen die leidende tegniese funksie en die kinematiese aspek – 'n analogie wat op die idee van

“konsekwente beweging” geskoei is. Hierdie is ’n eg kinematiese analogie in die tegniese, omdat die idee van konsistente beweging een is wat ons vir die eerste keer teëkom in daardie dimensie van ons ervaring. Ook geldigheid (as ’n meetbare eienskap) is ’n konstitutiewe konsep wat op die interne effek of krag van ’n tegniese artefak betrekking het en voortvloei uit die analogiese verband tussen die tegniese en die fisiese. Dit is in die fisiese dimensie dat die idee van krag, van oorsaak en gevolg, en van die werking van sake, oorspronklik voorkom; hier word dit in analogies tegniese sin aangewend. Hiervolgens kan ’n toets se geldigheid verdedig word wanneer die nodige tegniese effekte of uitkomst (toetsresultate) verkry word. In die skakeling van die tegniese funksie met die ruimtelike, biotiese en psigiese dimensies het die konsepte van beperkte omvang, differensiasie en intuïtiewe aanneemlikheid onderskeidelik hulle oorsprong.

’n Taaltoets, byvoorbeeld, het beperkte toepassingswaarde en kan nie vir doeleindes anders as waarvoor dit ontwerp is, aangewend word nie; daar is (tegniese) grense aan die vermoë van ’n toets (ruimtelike analogie). So ook moet ’n taaltoets inhoudelik gedifferensieer wees (biotiese analogie) en vir die gebruikers daarvan ’n gevoel van bevrediging gee, m.a.w. die gesigsgeldigheid daarvan moet deur ontwerpers verreken word (psigiese analogie). In die numeriese analogie is die idee van “technical systematicity” vervat. Hierdie konsep hou in dat die ontwerp van ’n toegepaste linguistiese artefak die sistematiese byeenbrenging en integrasie van ’n verskeidenheid tegniese oorwegings behels – ’n eenheid binne ’n veelheid van empiriese data en ontwerpbeginsels moet nagestreef word. Hierdie begrip speel veral in die subjektiewe proses van geldigmaking of validering ’n belangrike rol – bewyse vir die toereikendheid van ’n taaltoets kan wel gevarieer en uiteenlopend van aard wees, maar moet op ’n rasionele wyse, in die vorm van ’n tegniese argument, tot ’n eenheid geïntegreer kan word.

Regulatiewe idees ontstaan wanneer die leidende tegniese funksie ander modaliteite antisipeer of “vooruit oorweeg” en sodoende deur daaropvolgende modaliteite, spesifiek die linguale, sosiale, ekonomiese, estetiese, juridiese, etiese en geloofsdimensie, ontsluit word. Die noodigheid daarvan dat ’n linguistiese ontwerp in die vorm van ’n bloudruk of gedetailleerde plan (bv. toetsspesifikasies in die geval van ’n taaltoets) weergegee moet word, is ’n tegniese idee wat uit die linguale analogie resulteer. Die bloudruk is die tegniese uitdrukking van die plan, die verbeeldingryke ontwerp van ’n taaltoets. Ewesom die idee dat toetsresultate verkry moet word waarvan die betekenis tegnies geartikuleer (d.i. geïnterpreteer) kan word op ’n toepaslike en betekenisvolle manier (die interpretasie van toetsresultate is, soos uit die bespreking hierbo afgelei kon word, van kritiese belang in die hedendaagse beskouing van geldigheid). Uit die koppeling van die tegniese met onderskeidelik die sosiale, ekonomiese en estetiese dimensies ontspruit die analogiese momente van implementeerbaarheid, bruikbaarheid en tegniese harmonie of belyning tussen die verskeie komponente van ’n ontwerp. Deursigtigheid en verantwoordbaarheid is op hul beurt regulatiewe idees wat volg uit die verbinding van die tegniese met die juridiese, terwyl die vereiste van regverdigheid uit die etiese analogie spruit. Ten laaste hou pogings om ’n tegniese ontwerp vertrouens- en agtenswaardig te maak, met die geloofaspek daarvan verband. Die reputasie en geloofwaardigheid van ’n toets bring mee dat daar groter sekerheid en agting vir sy resultate is.



*Figuur 2.3 Konstitutiewe konsepte en regulatiewe idees in toegepaste linguistiese ontwerpe*

Weideman (2017) wys verder op die terugskouende en antisipatoriese aard van hierdie raamwerk. Dit behels dat die ontwerp van toegepaste linguistiese artefakte tegelyk vanuit twee perspektiewe of invalshoeke benader behoort te word. Besluite omtrent 'n ontwerp moet, in die eerste plek, retrospektief in terme van 'n tegniese-teoretiese basis verdedig kan word, terwyl daar terselfdertyd ook vooruitwerkend oor byvoorbeeld die moontlike (sosiale) gevolge van 'n ontwerp nagedink behoort te word. Hierdie raamwerk bring toetsontwerp dus nie net met intrinsieke, konvensionele ontwerpvoorwaardes (soos betroubaarheid en geldigheid) in verband nie, maar erken ook sosiale oorwegings (deursigtigheid, verantwoordbaarheid, regverdigheid) as inherente vereistes vir verantwoordelike toetsontwerp. Trouens, omdat beide konstitutiewe en regulatiewe ontwerpvoorwaardes in ag geneem word, blyk hierdie raamwerk 'n gebalanseerde middeweg tussen onderskeidelik die modernistiese (tradisionele) en postmodernistiese (hedendaagse) uitkyk op toegepaste linguistiek te vind. Weideman (2007a:599) skryf in hierdie verband dat:

In modernist approaches, the solution [to a language problem] is required to have both validity, and consistency or reliability, otherwise its authority and integrity are undermined. In postmodernist approaches, the solution when implemented must also have ethical dimensions, i.e. must be transparent, accountable, theoretically and politically defensible, and promote the interests of those affected by it.

Nog 'n belangrike kenmerk van hierdie raamwerk is die onderlinge verweefdheid van die verskillende analogiese konsepte, waardeur hulle gesamentlik 'n strukturele geheel van beginsels vir verantwoordelike toetsontwerp vorm. Van Dyk (2010:50-51) wys byvoorbeeld daarop hoedat die tegniese bruikbaarheid van 'n taaltoets (ekonomiese analogie) nie werklik sonder inagnome van die implementeerbaarheid (sosiale analogie) daarvan bepaal kan word nie. So ook maak konsepte soos betroubaarheid (kinematiese analogie) en geldigheid (fisiese analogie) normaalweg deel van die teoretiese begroning (analitiese analogie) van 'n toets uit.

Laastens beklemtoon Weideman (2017) dat daar altyd 'n sobere balans tussen die konstitutiewe en regulatiewe momente van 'n ontwerp gehandhaaf behoort te word. Verantwoordelike toetsontwerp kom in hierdie opsig op die uitvoer van 'n



versigtige balanseertoertjie tussen 'n verskeidenheid van (dikwels botsende) ontwerpvereistes neer, soos Weideman (2006:83) tereg opmerk: “the test designer should carefully weigh a variety of potentially conflicting demands, and opt not only for the socially most appropriate, but also for a frugal solution”.

Laasgenoemde impliseer dat toegepaste linguïste bereid sal moet wees om weloorwoë en redelike toegewings te maak wanneer hulle toetse ontwerp. So byvoorbeeld kan logistieke en administratiewe beperkinge (sosiale analogie) wat 'n nadelige invloed op die bruikbaarheid van 'n toets (ekonomiese analogie) kan hê, daartoe lei dat ontwerpers uiteindelik met 'n laer (maar steeds aanvaarbare) betroubaarheidswaarde (kinematiese analogie) tevrede sal moet wees. Wat spesifiek die onderhawige akademiese geletterdheidstoets betref, word daar in latere hoofstukke aangedui in hoe 'n mate sodanige kompromieë deel van hierdie toets se ontwerpproses uitgemaak het.

Ten slotte word 'n skematiese opsomming van die raamwerk wat hierbo bespreek is, verskaf (kyk tabel 2.7), soos oorgeneem en vertaal uit Weideman (2017:224). Daarnaas volg 'n opgaaf van ontwerpbeginsels (kyk tabel 2.8) wat uit hierdie raamwerk voortspruit en wat, in elke geval, dienooreenkomstig met 'n bepaalde terugskouende en/of antisipatoriese analogiese moment in die laaste kolom van die onderstaande tabel korrespondeer.

<b>Toegepaste linguistiese ontwerp</b>	<b>Aspek / funksie / dimensie / modaliteit</b>	<b>Tipe funksie</b>	<b>Terugskouende / antisipatoriese analogiese moment</b>
word gebaseer op	numeriese	konstitutiewe	eenheid binne 'n veelheid van ontwerpbeginsels / bronne van bewyslewering
	ruimtelike		beperkinge, omvang
	kinematiese		interne konsekwentheid (tegniese betroubaarheid)
	fisiese		interne effek / krag (geldigheid)
	biotiese, organiese		differensiasie
	psigiese		intuïtiewe aanneemlikheid (gesigsgeldigheid)
	analitiese	funderende	rasionaal vir ontwerp
word gekwalifiseer deur	tegniese	kwalifiserende / leidende funksie (van die ontwerp)	
word ontsluit deur	linguale	regulatiewe	artikulasie van ontwerp by wyse van bloudruk / kurrikulum / plan
	sosiale		implementering / administrasie
	ekonomiese		tegniese bruikbaarheid of nut, spaarsaamheid
	estetiese		harmonisering van konflikte, regstel van belyningsfoute
	juridiese		deursigtigheid, verdedigbaarheid, regverdigheid, legitimiteit
	etiese		verantwoordbaarheid, sorgsaamheid, diens
	geloof		agting en vertrou

*Tabel 2.7 Konstitutiewe en regulatiewe momente in toegepaste linguistiese ontwerpe*

1.	Beoordeel 'n toets as 'n geheel met geïntegreerde dele. Integreer, op 'n sistematiese manier, veelvuldige bewyse in die formulering van 'n geldigheidsargument vir 'n toets as geheel en vir sy dele (subtoetse) en eienskappe.
2.	Stip duidelik aan en verduidelik aan die publiek die toepaslik beperkte reikwydte van die toets, asook die grense en beperkings van sy meting, en doen dit op 'n beskeie manier.
3.	Verseker dat die toetsresultate wat verkry word, toereikend konsekwent is, ook met die verloop van tyd. Sien toe dat die toets betroubaar meet, ook oor verskeie aanwendings daarvan.
4.	Verseker doeltreffende meting deur van 'n meetinstrument gebruik te maak waarvan die toereikendheid (geldigheid) verdedig kan word. Sien toe dat die toetsresultate effektief verkry word deur die meting.
5.	Gebruik 'n toets wat toepaslik en toereikend gedifferensieerd is, veral wat betref die tipe en verskeidenheid subtoetse waarmee die konstruk op 'n gepaste manier tot sy reg kom in die meting.
6.	Maak die toets intuïtief aanneemlik en aanvaarbaar vir die gebruikers daarvan (oorweeg die gesigsgeldigheid van 'n toets).
7.	Verskaf in die mees resente terme 'n teoretiese verdediging van dit wat getoets word. Bied 'n afdoende teoretiese rasionaal aan vir die toetskonstruk.
8.	Bring die plan en ontwerp van die toets na vore deur dit uit te druk in 'n bloudruk, wat die konstruk operasionaliseer en die toetsspesifikasies behoorlik uitstpel. Maak seker dat die toets interpreteerbare en betekenisvolle resultate oplewer.
9.	Maak die toets, asook inligting omtrent die toets, toeganklik vir almal. Ont sien geen moeite om inligting aangaande die toets, en ook hoe en vir welke doeleindes dit saamgestel en ontwerp is, openbaar te maak nie. Verseker dat die aanwending van die toets sosiaal toepaslik is.
10.	Verkry die toetsresultate op 'n effektiewe manier en verseker dat dit bruikbaar is. Weeg verskillende moontlikhede teen mekaar af ten einde nuttigheid en doeltreffendheid in balans met mekaar te bring.
11.	Belyn die toets met die onderrig wat óf daarop volg óf dit voorafgaan, en so na as moontlik met die leerproses. Onthou om ook belyning van die toets met die onderrigprosesse wat dit voorafgaan of daarop volg, te verseker, asook om dit te harmoniseer met die taalbeleid wat die meting vereis.
12.	Wees bereid om aan die publiek verslag te doen oor hoe die toets gebruik is of gebruik gaan word. Laat geen steen onaangeroerd om nie net aan eweknieë nie, maar ook aan die publiek te verduidelik hoekom die toets aangewend moet word.
13.	Heg waarde aan die integriteit en kwaliteit van die toets. Maak geen toegewings wat die toets se status as regverdige meetinstrument sal ondermyn nie. Verseker dat die toets sover moontlik die wat dit moet aflê, ten goede raak.
14.	Ont sien geen moeite om die toets toepaslik vertrouens- en agtenswaardig te maak nie. Die geloofwaardigheid van die toets is 'n voorwaarde vir die reputasie wat dit mettertyd sal verkry uit al die sorgvuldige ontwerp en regverdige aanwending daarvan.

(Oorgeneem, vertaal, aangevul en aangepas uit Weideman, 2012:8-9)

*Tabel 2.8 Beginsels vir verantwoordelike toetsontwerp*

#### 4. Slot

Die ontwerpbeginsels in tabel 2.8 tesame met die skema in tabel 2.7 vorm die teoretiese ruggraat van hierdie studie deurdat dit, soos uit die daaropvolgende hoofstukke sal blyk, deurlopend as ontwerpmaatstaf geld, en die ontwerp van die betrokke Afrikaanse nagraadse toets van akademiese geletterdheid ten grondslag lê. Die doel van hierdie hoofstuk was immers om aan die hand van 'n verskeidenheid bronne oor die toegepaste linguistiek en taaltoetsing 'n grondige teoretiese basis aan hierdie studie te verskaf. In 'n studie soos die een voorhande word dit noodsaaklik geag dat die ontwerp van 'n meetinstrument – 'n saak wat pertinent vanaf die volgende hoofstuk aan die orde kom – eers deur behoorlike teoretiese nadenke voorafgegaan moet word. Bachman (2004:4) konstateer in hierdie verband dat “applied linguistic theory [...] guides the design of assessment tasks, as we attempt to develop test tasks that will reflect language use outside of the test itself and that will engage the abilities we want to assess”.

Hierdie siening strook volkome met dié van Weideman (2007) wat, soos in die inleiding van hierdie hoofstuk vermeld, vertroutheid met die teoretiese grondbeginsels van die toegepaste linguistiek as 'n onmisbare voorvereiste vir verantwoordelike toetsontwerp stel. Die gemis aan 'n teorie van die toegepaste linguistiek kan dit wat Bachman (2004) hier voorstaan, belemmer, en die uiteensetting hierbo van Weideman (2012, 2017) se voorstel om aan so 'n teorie gestalte te begin gee, is dus 'n poging om een moontlike weergawe van so 'n teorie te beproef in die teoretiese en praktiese regverdiging van die ontwerp van een so 'n toegepaste linguistiese artefak, TAGNaS, wat die fokus van hierdie studie uitmaak.

## **Hoofstuk 3**

### **Die ontwerp van 'n nagraadse toets van akademiese geletterdheid in Afrikaans**

#### **1. Inleiding**

Hierdie hoofstuk, tesame met die daaropvolgende hoofstukke in hierdie studie, dokumenteer die ontwikkelingsverloop van die Toets van Akademiese Geletterdheid vir Nagraadse Studente (TAGNaS) vanaf die konseptualisering, ontwerp en loodstoetsing daarvan tot die ontleding van toetsresultate en die gepaardgaande evaluering van items vir moontlike implementering. Empiriese data wat tydens hierdie proses versamel is, word gebruik om bepaalde aannames en gevolgtrekkings omtrent die konseptuële toets te maak. Verder geskied die dokumentering van gegewens oor die ontwerp en verfyning van TAGNaS deurgaans met die inagneming van die raamwerk van ontwerpbeginsels vir verantwoordelike toetsontwerp wat in Hoofstuk 2 breedvoerig behandel is.

Die optekening en boekstaaf van die ontwikkelingsgang van so 'n meetinstrument vertel in die woorde van Shohamy (2001) “die storie van 'n toets” – 'n nosie van wat in noue verband staan met die regulatiewe ontwerpvoorwaardes van deursigtigheid en verantwoordbaarheid, twee analogiese momente wat voortvloei uit die skakeling van die leidende tegniese funksie van sulke ontwerpe met onderskeidelik die juridiese en etiese dimensies van die werklikheid. Op hierdie manier word inligting omtrent die onderhawige toets openbaar en toeganklik gemaak vir alle belanghebbendes en belangstellendes en dien dit as bewys daarvan dat waarde aan die kwaliteit van die ontwerpte items geheg word. Van der Slik en Weideman (2008:363) beskou sodanige onderneming as “a necessary first step in the process of becoming transparent, and subsequently, accountable as test developers”.

Meer spesifiek nog val die fokus in Hoofstuk 3 op die konstruk van akademiese geletterdheid waarop die ontwerp van TAGNaS gebaseer word. Een van die eerste belangrike stappe in die toetsontwerpsiklus (vgl. Fulcher, 2010:93-96) behels die vind van 'n gepaste konstruk wat as basis vir die ontwerp kan dien. Hierdie hoofstuk bied 'n beskrywing van die konstruk wat die teoretiese rasionaal vir dit wat in hierdie studie getoets word, verskaf. Die artikulasie van 'n teoretiese rasionaal is in die eerste plek 'n ontwerpvoorwaarde wat duidelik uit die analitiese funderingsfunksie van 'n linguistiese ontwerp resulteer. As een van 'n veelheid van bronne van bewyslewering maak hierdie artikulasie egter ook 'n waardevolle bydrae tot 'n geïntegreerde argument vir die konstrukgeldigheid van TAGNaS, waaruit die numeriese analogieë – van 'n tegniese eenheid en veelheid, asook die fisiese analogie van tegniese geldigheid – duidelik blyk. Verder gee dit 'n aanduiding van die beperkte metingsomvang van hierdie toets ('n ruimtelike analogie binne die tegniese sfeer), naamlik dat die konstruk wat gemeet word spesifiek een van akademiese geletterdheid is, en nie bv. taalvermoë in die algemeen nie. Davies, Brown, Elder, Hill, Lumley & McNamara (1999:207) definieer die konstrukgeldigheid van 'n taaltoets as “an indication of how representative it is of an underlying theory of language learning”. Konstrukvalidasie behels daarom 'n ondersoek na die kwaliteite wat deur 'n toets gemeet word om sodoende gronde vir die rasionaal daarvan te verskaf, en dit dan in 'n argument te kan integreer en aan te bied.

Die operasionalisering van die konstruk in die vorm van toetsspesifikasies en toepaslike taaktipes is die volgende aspek waaraan aandag gegee word. Hierdie hoofstuk bevat 'n volledige bespreking van die subtoetse wat vir TAGNaS beplan word, asook aanduidings van hoe hierdie subtoetse waarvan sommige gewysigde of nuwe subtoetse is, aansluiting vind by die toetskonstruk, en in die besonder ook by die analitiese karakter van akademiese diskoers wat, soos spoedig sal blyk, deur sommige as die wesenskenmerk van 'n konstruk van akademiese geletterdheid gereken word. Meer inligting word ook voorsien rakende die meerfasige toetsmodel wat potensieel vir TAGNaS in die vooruitsig

gestel word. Verdere ontwerpaspekte waarvoor motivering in hierdie hoofstuk verskaf word, sluit onder andere in: die gebruik van die veelkeuseformaat, die inkorporering van 'n akademiese skryftaak in die toetsontwerp, die keuse van 'n geskikte toetsstema en leestekste, asook die basering van die woordeskatitems se ontwerp op die nuut ontwikkelde Akademiese Woordelys vir Afrikaans soos saamgestel deur Steyn (2013).

## **2. Die keuse van 'n toetskonstruk: 'n Hersiene konstruk van akademiese geletterdheid as teoretiese rasionaal**

### **2.1 Die toetskonstruk van TALL, TAG en TALPS**

Alvorens toetsitems ontwerp kan word, moet daar in duidelike en meetbare terme uitgestippel word watter vermoë(ns) gemeet word. Bachman en Palmer (2010:43) omskryf 'n toetskonstruk as 'n “specific definition of an ability that provides the basis for a given assessment or assessment task and for interpreting scores derived from this task”. Soos reeds kortliks in Hoofstuk 1 vermeld is, is die besluit geneem om TAGNaS te grond op 'n onlangs-geamendeerde weergawe van dieselfde konstruk wat as basis vir die ontwikkeling van die Test of Academic Literacy Levels (TALL), Toets van Akademiese Geletterdheid (TAG) en Test of Academic Literacy for Postgraduate Students (TALPS) gebruik is. Die oorspronklike konstruk wat in 2003 sy beslag gekry het, het ontstaan nadat die Eenheid vir Akademiese Geletterdheid aan die Universiteit van Pretoria (UP) die geskiktheid van die English Literacy Skills Assessment for Tertiary Education (ELSA PLUS) wat sedert 2000 gebruik is om eerstejaarstudente se taalvaardigheid te meet, begin bevraagteken het. Die ELSA PLUS wat in samewerking met die Hough en Horne konsultasiefirma ontwikkel is, het uit sewe afdelings bestaan, naamlik (Van Dyk & Weideman, 2004a):

- klankleer (die vermoë om tussen klanke te onderskei en dit te herken);
- diktee (die vermoë om gesproke taal neer te skryf en jou aan akademiese skryfkonvensies te hou);

- basiese gesyferdheid (taalgebruik word geïntegreer met 'n basiese vertroudheid met syfers);
- leesbegrip (op elementêre en intermediêre vlak);
- taal en grammatika van ruimteverhoudinge;
- 'n sinsluitingstoets (die vermoë om 'n betekenisvolle geheel te vorm deur sinne op 'n samehangende manier te voltooi);
- woordeskat binne konteks (die vermoë om relevante inligting uit 'n bepaalde konteks te haal ten einde die betekenis van sekere woorde of frases te bepaal).

Van Dyk en Weideman (2004a:5,7) wys daarop dat 'n toetskonstruk gewoonlik in terme van 'n (taal)teorie tot uiting kom, d.i. ter staving of bevestiging daarvan. Definisies van taalvermoë hang uiteindelik af van die siening van taal wat deur die betrokke toetsontwerper(s) gehuldig word. Die beskrywing hierbo onthul duidelik hoedat die ELSA PLUS van die teoretiese standpunt uitgaan dat (akademiese) taalvermoë in terme van die strukturele komponente daarvan gedefinieer en gemeet behoort te word of behoort te kan word. Volgens hierdie sienswyse is taal slegs 'n kombinasie van klank, vorm en betekenis. Omrede hierdie beskouing omtrent taal (en taaltoetsing) as verouderd en beperkend geag word (Van Dyk & Weideman, 2004a:4) (vgl. die vorige bespreking van die psigometriese benadering en diskrete itemtoetsing *teenoor* kommunikatiewe toetsing), is die besluit destyds geneem om 'n nuwe, meer geskikte konstruk van akademiese geletterdheid in die plek van die ELSA PLUS-konstruk te ontwikkel.

Van Dyk en Weideman (2004a:5) motiveer dié oorskakeling soos volg:

For more than thirty years now, it has been accepted that the design of a language test depends on a broader, richer perspective on language than the views held earlier. This is a perspective that goes beyond a restrictive view of language that limits it to a combination of sound, form and meaning, or, in technical linguistic terms, phonological, morphological, syntactic and semantic elements. In line with socially enriched views of language, such a broader framework for language maintains that language is not only expressive, but communicative, intended to mediate and negotiate human interaction.



Die verskille tussen 'n beperkende en oop siening van taal, en die implikasies wat dit vir taaltoetsing inhou, kan soos volg saamgevat word (Weideman, 2003:58):

Beperkend	Oop
Taal bestaan uit hierdie elemente: <ul style="list-style-type: none"> <li>• klank</li> <li>• vorm, grammatika</li> <li>• betekenis</li> </ul>	Taal is 'n sosiale instrument om: <ul style="list-style-type: none"> <li>• menslike interaksie te bewerkstellig</li> <li>• en te bemiddel</li> <li>• in spesifieke kontekste</li> </ul>
<i>Hooffunksie</i> : uitdrukking	<i>Hooffunksie</i> : kommunikasie
<i>Aanleer van taal</i> = bemeestering van struktuur	<i>Aanleer van taal</i> = bedreweheid in kommunikasie
<i>Fokus</i> : taal as objek	<i>Fokus</i> : proses van taalgebruik

*Tabel 3.1 Twee sieninge van taal*

'n Bykomende, tog verwante, punt van kritiek teen die teoretiese onderbou van die ELSA PLUS is dat dit taalvermoë beskou as bestaande uit 'n hele paar afsonderlike vaardighede, in hierdie geval luister-, lees- en skryfvaardigheid. Hierdie toets volg dus die tradisionele “skills-based approach to language ability” (Van Dyk & Weideman, 2004a:6-7). Hoewel baie invloedryk in taaltoetsing sedert die tweede helfte van die vorige eeu, staan verskeie taaltoetsteoretici krities teenoor hierdie benadering (vgl. Bachman & Palmer, 2010; Fulcher & Davidson, 2007; Douglas, 2010). In die eerste plek gee dit oorsprong aan die aanname dat hierdie vaardighede in 'n toetssituasie geïsoleerd van mekaar gemeet en bepunt kan word, terwyl in die meeste gevalle

[..] we combine these skills in communicative language use: we read a chemistry textbook, taking notes for later use in oral discussion; we engage in rapid listening/speaking give and take in a telephone conversation; we read an email message and immediately compose a reply [...] (Douglas, 2010:19)

In aansluiting hierby is daar die verdere beperking dat hierdie benadering take of aktiwiteite wat taal op heeltemal uiteenlopende maniere aanwend, onder 'n enkele vaardigheid saamtrek. So byvoorbeeld verskil 'n aangesig-tot-aangesig-gesprek aansienlik van 'n radionuusuitsending, hoewel beide aktiwiteite luistervaardigheid betrek. Die vaardighedsgebaseerde benadering laat ook na om in aanmerking te neem dat taalgebruik nie bloot 'n algemene verskynsel is wat in 'n vakuum plaasvind nie. Bachman en Palmer (2010:55) stel dit dat “[w]e

do not just ‘read’; we read about something specific, for some particular purpose, in a particular setting”. Volgens hierdie outeurs vind taalgebruik plaas, of verweselik taalvermoë sigself, in die uitvoering van spesifieke take waarin taal doelbewus, of met “communicative intent” (Douglas, 2010:19), gebruik word. Vir hulle is ’n vaardigheid, soos luister of lees, ’n gekontekstualiseerde manifestasie van die vermoë om taal in die uitvoering van ’n spesifieke taak of aktiwiteit te gebruik. Hulle argumenteer daarom dat “it is not useful to think in terms of ‘skills’, but rather to think in terms of specific activities or tasks in which language is used purposefully” (Bachman & Palmer, 2010:56). Eerder as om ’n vaardigheid in abstrakte terme te probeer omskryf, moet dit aan die taalgebruik van ’n spesifieke taak of aktiwiteit, gekoppel word, en dienooreenkomstig moet die karaktereenskappe daarvan, asook die komponente van taalvermoë wat daarby betrek word, beskryf word. Van Dyk (2010:16) merk in hierdie verband op dat “’n werklike toets van akademiese geletterdheid dus eerder die meting van studente se kommunikatiewe vermoë binne akademiese verband [is] as ’n toets wat slegs vaardigheidsvlakke meet”. Om sodanige toets die lig te laat sien, glo Van Dyk en Weideman (2004a:6) dat dit nodig is om te “think of academic language proficiency in terms of academic discourse, and conceive of the problem of measuring that ability as a problem of measuring the level to which academic discourse has been acquired”.

Die vraag wat vervolgens ontstaan het, was hóé ’n konstruk van akademiese geletterdheid, gebaseer op ’n oop, verruimde siening van taal en akademiese taalvermoë (kyk weer tabel 3.1) waarskynlik dan sou lyk. Om ’n antwoord hierop te vind, het die ontwikkelaars van die voorgestelde, nuwe konstruk verskeie definisies van (akademiese) taalvermoë geraadpleeg en oorweeg, spesifiek dié van Blanton (1994) en Bachman en Palmer (1996), tesame met Nan Yeld (2000) en haar kollegas aan die Universiteit van Kaapstad se herinterpretasie van laasgenoemde. Volgens Van Dyk en Weideman (2004a:7-11) het al drie hierdie definisies sigself op meer as een manier as bruikbaar in die ontwikkeling van ’n teoreties meer regverdigbare konstruk bewys, soos hieronder aangetoon word.

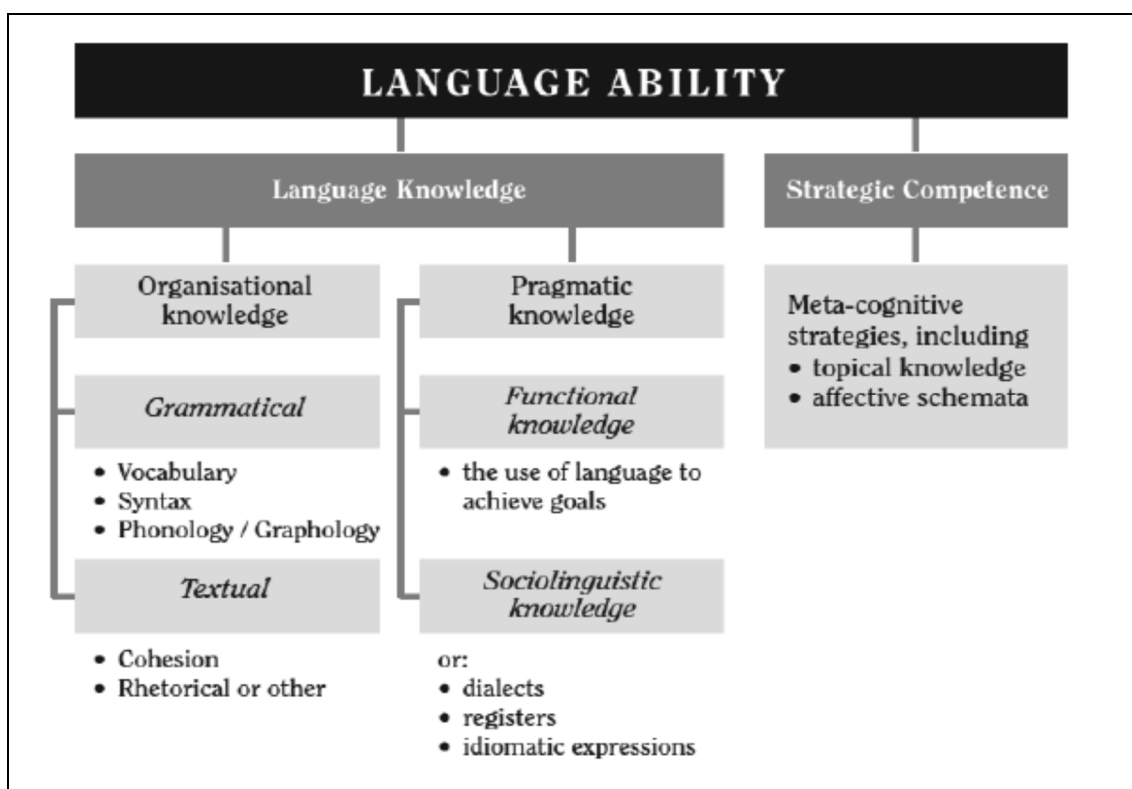
In die eerste plek is Blanton (1994:6) van mening dat om vanuit 'n sosiale, kommunikatiewe oogpunt by 'n funksionele definisie van akademiese diskoers uit te kom, dit nodig is om die regte vrae te vra. In plaas daarvan om te vra *wat akademiese diskoers is*, en dit dan in terme van linguistiese kenmerke en tipiese vorme te omskryf, moet daar eerder ondersoek ingestel word na *wat akademiese lesers en skrywers doen*. Sodra aanvaar word dat taal 'n sosiale aktiwiteit en medium is wat mense in staat stel om handelend (teenoor en in interaksie met mekaar) op te tree, behoort die fokus op aksies en gedrag te val, en nie bloot op kennis en linguistiese vorm nie. Blanton (1994:7) voer aan dat “[b]y focusing on behaviours, language takes its rightful place as a medium and a means of executing certain tasks, formed and shaped by the need to complete the task at hand. This is real language”. Let terloops hier op hoedat Blanton, deur op gedrag te fokus, in aansluiting by Bachman en Palmer (2010), taalgebruik in terme van doelbewuste taakuitvoering konseptualiseer. Blanton (1994:8) beskou interaksie met tekste as die kern van akademiese lees- en skryfgedrag. Volgens haar moet akademiese lesers en skrywers die volgende take of aksies suksesvol kan uitvoer:

1. Interpreteer tekste in die lig van eie ervaring en interpreteer eie ervaring in die lig van tekste;
2. Stem saam met tekste of verskil daarvan in die lig van daardie ervaring;
3. Bring tekste met mekaar in verband;
4. Sintetiseer tekste en gebruik hierdie samevoeging om nuwe idees te formuleer;
5. Ekstrapoleer uit tekste;
6. Skep eie tekste deur enige van, of al, die bogenoemde te doen;
7. Skryf en praat daaroor om enige van, of al, die bogenoemde te doen;
8. Voer punt 6 en 7 sodanig uit dat daar aan die verwagtinge van die teikengehoor voldoen word.

Die waarde van Blanton (1994:8) se definisie is daarin geleë dat dit afstand doen van die eng opvatting dat akademiese taalverwerwing met die bemeestering van bloot woordeskat en grammatikale konvensies in verband gebring kan word. Van

Dyk en Weideman (2004a:8) beskryf Blanton se omskrywing as ruimer en meer inklusief, aangesien plek gemaak word vir taalfunksies en vermoëns soos ‘eie akademiese stem’, ekstrapolering, sintetisering en argumentering, hetsy spesifiek vermeld of geïmpliseer. Bevoegdhede wat, by nadenke, tog seer seker in ’n konstruk van akademiese geletterdheid neerslag behoort te vind.

Voorts verskaf Bachman en Palmer (1996) ’n nóg meer gedetailleerde definisie van taalvermoë wat skematies soos volg voorgestel kan word (sien Van Dyk & Weideman, 2004a:8):



Tabel 3.2 Bachman en Palmer se konstruk

Hierdie outeurs bied ’n verklaring van taalvermoë binne ’n interaksionele raamwerk van taalgebruik, met die uitgangspunt dat

[I]language use involves complex and multiple interactions among the various individual characteristics of language users, on the one hand, and between these characteristics and the characteristics of the language use or testing situation, on the other. (Bachman & Palmer, 1996:62)

Gegronde op ’n vroeëre model van Bachman (1990) definieer Bachman en Palmer (1996:66-70) hier taalvermoë as bestaande uit twee hoofkomponente, naamlik

taalkennis ('language knowledge') en strategiese bekwaamheid ('strategic competence'), waar laasgenoemde na metakognitiewe strategieë, soos aktuele kennis en affektiewe skemata, verwys. Ter beskikking van hierdie strategieë is daar taalkennis wat as 'n inligtingsdomein van onderskeidelik organisatoriese en pragmatiese kennis in die geheue beskryf kan word. Organisatoriese kennis stel 'n persoon in staat om taaluitinge op grammatikale en tekstuele vlak te organiseer, terwyl pragmatiese kennis dit moontlik maak om die bedoelinge van taalgebruikers en die verbande tussen taaluitinge te interpreteer (funksionele kennis), asook om geskikte taal vir 'n bepaalde taalgebruikskonteks voort te bring en te interpreteer (sosiolinguistiese kennis). Volgens hierdie outeurs lê die vermoë om diskoers te skep en te vertolk dus opgesluit in 'n kombinasie van taalkennis en metakognitiewe strategieë.

Hoewel Bachman en Palmer (1996) se klassifikasie tekortkominge blyk te hê – Van Dyk en Weideman (2004a:8-9) dui byvoorbeeld aan hoedat sekere van die kategorieë gedeeltelik oorvleuel – maak dit die raamwerk nie noodwendig minder toepasbaar nie, veral in die lig van hulle opmerking dat hierdie model alleenlik as 'n breë rigtingwyser vir ontwikkelaars van taaltoetskonstrukte beskou behoort te word. Hulle wys daarop dat

[i]t is not our intention to suggest that all language testing should be based on all, or even specific parts of this particular model. At the same time, however, we believe very strongly that the consideration of language ability in its totality needs to inform the development and use of any language test. (Bachman & Palmer, 1996:67)

Hulle beklemtoon ook die noodsaaklikheid daarvan om hierdie raamwerk vir elke spesifieke taalgebruiks- of toetskonteks wat ter sprake mag kom, te herinterpreteer en kontekstueel aan te pas:

We also need to define language ability in a way that is appropriate for each particular testing situation [...] The way we define language ability for a particular testing situation, then, becomes the basis for the kinds of inferences we can make from the test performance. (Bachman & Palmer, 1996:66)

Die primêre doel van taaltoetse is om gevolgtrekkings oor kandidate se taalvermoë te maak. Hierdie interpretasies moet bo en behalwe die spesifieke take wat geassesseer word, kan veralgemeen, hoewel dit tewens ook nie vir enige

of *alle* taalgebruikstake of -kontekste kan geld nie. Die interpretasies is veronderstel om nie verder te strek of te veralgemeen as die spesifieke taalgebruikstake waarmee kandidate heel waarskynlik in 'n bepaalde konteks of domein te doen gaan kry nie (Bachman & Palmer, 2010:60).

Die tersaaklike domein in die geval van 'n konstruksie van akademiese geletterdheid is die hoërsonderwyskonteks. Hiermee in gedagte het die ontwikkelaars van die nuwe konstruksie hulle ten laaste gewend tot die werk van die Alternative Admissions Research Project (AARP) aan die Universiteit van Kaapstad waaruit die ontwikkeling van die Placement Test in English for Educational Purposes (PTEEP) voortgevloei het (Yeld, 2000). As deel van AARP se navorsing is die oorspronklike Bachman-en-Palmer-konstruksie, soos hierbo uiteengesit, vir die Suid-Afrikaanse hoërsonderwysomgewing hervertolk. Vir hierdie herinterpretasie is bydraes van kundige panele, spesifiek in die vorm van insigte omtrent tipiese akademiese take, by die oorspronklike Bachman-Palmer-indeling ingelyf. Dit het die identifikasie van verskeie taalfunksies en akademiese geletterdheidstake soos teksbegrip, parafrasering, opsomming, beskrywing, argumentering, klassifisering, kategorisering, vergelyking en kontrastering ingesluit (Van Dyk & Weideman, 2004a:9).

Verskille tussen die AARP-toetsontwerp en die beoogde nuwe toets van die UP het egter die ontwikkelaars van laasgenoemde verhinder om die bogenoemde AARP-toetskonstruksie net so te gebruik. Die verskille het spesifiek betrekking op die doel en logistieke haalbaarheid van die toetse, asook met die gedifferensieerdheid van die subtoetse. Om mee te begin, was die AARP-toets 'n *toelatingstoets* waarvan die resultate 'n aanduiding van die akademiese potensiaal (dus ook kognitiewe vermoë) van kandidate moes gee, terwyl dit by die UP gegaan het om 'n *plasingstoets* van akademiese geletterdheid op grond waarvan reeds toegelate studente se taalrisikolakke bepaal moes word vir plasing in toepaslike ondersteuningskursusse. Boonop was die AARP-toetsontwerp moeilik om te administreer, veral in gevalle waar groot groepe

studente binne die bestek van 'n enkele dag getoets moes word, soos wat by die UP die geval was, met in daardie stadium tot soveel as 6000 studente. Die besluit is gevolglik geneem om op basis van die AARP-konstruk met 'n vereenvoudigde, geherformuleerde bloudruk vorendag te kom wat betroubare akademiese geletterdheidstoetsing binne veel strenger tydsbeperkinge moontlik sou maak (Van Dyk en Weideman, 2004a:10).

Die voorgestelde nuwe konstruk wat in die plek van die ELSA PLUS-konstruk ontwikkel is, en waarop die ontwerp van TALL, TAG en TALPS uiteindelik gebaseer is, vereis dat studente daartoe in staat moet wees om (Van Dyk & Weideman, 2004a:10):

- 'n reeks akademiese woordeskat in konteks te verstaan;
- metafore en idioome te interpreteer en te gebruik, asook konnotasie, woordspeling en dubbelsinnigheid raak te sien en te begryp;
- verbande tussen verskillende dele van 'n teks te verstaan, bewus te wees van die logiese ontwikkeling van 'n (akademiese) teks vanaf inleiding tot gevolgtrekking, en te weet hoe om taal te gebruik wat die verskillende dele van 'n teks laat saamhang;
- verskillende tekstipes (genres) te interpreteer en begrip te toon vir die betekenis wat daardeur oorgedra word en die gehoor op wie dit gemik is;
- inligting in grafiese of visuele formaat te interpreteer, te gebruik en te skep;
- onderskeidings te maak tussen noodsaaklike en nienoodsaaklike inligting, feite en menings, stellings en argumente en om onderskeid te tref tussen oorsaak en gevolg, om data wat vergelykings tref te klassifiseer, kategoriseer en hanteer;
- opeenvolging en volgorde raak te sien, om eenvoudige numeriese skattings en berekeninge te doen wat op akademiese inligting betrekking het, die tref van vergelykings toelaat en vir argumentasiedoeleindes gebruik kan word;

- te weet wat as bewyse vir 'n argument geld, om uit inligting te ekstrapoleer deur afleidings te maak en die inligting of die implikasies daarvan op ander gevalle as die een voorhande toe te pas;
- die kommunikatiewe funksie van verskeie uitdrukkingswyses in akademiese taal te verstaan (soos definiëring, die verskaffing van voorbeelde, argumentering); en
- betekenis te skep (bv. van 'n akademiese teks) verby sinsvlak.

Tot op datum kan heelparty bewyse ter ondersteuning van die bostaande konstruk se teoretiese gegrondheid en geldigheid aangebied word. TALL, TAG en TALPS het sigself deur die jare heen as hoogs betroubare en bruikbare meetinstrumente gevestig, met vele publikasies en studies (o.a. die doktorsale studies van Van Dyk, 2010 en Rambiritch, 2012) ter staving. Boonop is die konstruk sedert sy konseptualisering by verskeie konferensie- en seminaargeleenthede voorgelê waar dit oor die algemeen wye en positiewe reaksie uitgelok het (Van Dyk & Weideman, 2004a:11). Agterna beskou, blyk die ontwikkelaars se aanvanklike oortuiging dat “the abilities described [in the construct] echo strongly [...] what it is that students are required to do at tertiary level” (Van Dyk & Weideman, 2004a:11) dus in groot mate positief bevestig te wees.

Ondanks die onmiskenbare en bewese geslaagdheid van hierdie toetskonstruk neem Patterson en Weideman (2013b) die inhoud daarvan tog onlangs opnuut in hersiening. Hulle beskou sodanige herbesoek as tydig en geleë, aangesien geen kritiese ondersoek van hierdie konstruk wat reeds meer as 'n dekade lank in gebruik is, tot dusver onderneem is nie. Omdat 'n konstruk so 'n belangrike onderdeel van 'n taaltoets uitmaak, is toetsontwikkelaars “[...] constantly seeking to improve their theoretical understanding of what it is that they seek to measure” (Patterson & Weideman, 2013a:108). In die afdeling wat volg, word Patterson en Weideman (2013b) se voorgestelde verfynde weergawe van die oorspronklike konstruk volledig uiteengesit, met 'n beknopte oorsig van die teoretiese



beweegredes vir hierdie verfyning. Hierdie uiteensetting is van oorwegende belang, aangesien dit uitdrukking gee aan die teoretiese rasionaal waarop die ontwerp van TAGNaS gebaseer word. Die feit dat 'n bygewerkte definisie van akademiese geletterdheid as fundering gebruik word, is 'n element van vernuwing waarvan reeds in Hoofstuk 1 melding gemaak is.

## 2.2 'n Verfynde konstruk as basis vir die ontwerp van TAGNaS

Soos reeds duidelik gemaak is, kan 'n toetskonstruk omskryf word as 'n definisie van 'n sekere vermoë of vermoëns vir metingsdoeleindes. Sodanige definiëring geskied gewoonlik vanuit 'n bepaalde perspektief. In die geval van konstrukte van akademiese geletterdheid – waar dit spesifiek om die hantering en beheersing van akademiese diskoers gaan – word daar noodwendig, volgens Patterson en Weideman (2013a:107), op definisies gesteun wat aanneem dat akademiese diskoers 'n eiesoortige tipe diskoers is wat van ander soorte diskoers verskil. Hulle gaan gevolglik van die veronderstelling uit dat “if we improve our understanding of what makes academic discourse unique, we can therefore potentially improve our test designs” (Patterson & Weideman, 2013a:107). Hiervolgens behoort 'n konstruk van akademiese geletterdheid dus sy oorsprong te vind in, en ook verfyn te word vanuit, 'n idee omtrent die eiesoortige aard van akademiese diskoers. In aansluiting hierby voer Bachman en Palmer (1996:24) aan dat om die outentisiteit van 'n toets te verseker “we first [need to] identify the critical features that define tasks in the TLU [Target Language Use] domain”.

Patterson en Weideman (2013b) begin hulle betoog deur verskeie bestaande definisies van akademiese diskoers te problematiseer. Hulle toon deur 'n gedetailleerde kritiese ontleding aan hoedat die definisies van teoretici soos Flower (1990), Suomela-Salmi en Dervin (2009), Livnat (2012) Gunnarsson (2009), Hyland (2011) en Snow en Uccelli (2009) nie volkome daarin slaag om 'n tiperende kenmerk te identifiseer wat die eiesoortigheid van akademiese diskoers sekuur vasvang en dit van ander diskoerstipes onderskei nie. In die

meeste gevalle kan die kenmerke van akademiese diskoers wat uitgelig word (bv. dialogiese aard, oorredingskrag, die klem op feitelikheid, ens.) ook op ander soorte diskoers van toepassing gemaak word. Die groot probleem blyk, in die woorde van Snow en Uccelli (2009), te wees dat “dozens of traits have been identified [...] that might function as markers of academic language, but it is unclear that any of them actually defines the phenomenon” (aangehaal in Patterson & Weideman, 2013b:131).

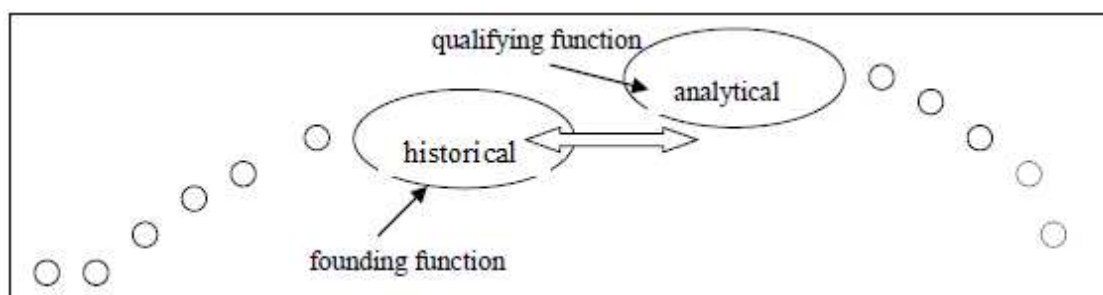
As ’n moontlike oplossing vir hierdie probleem stel Patterson en Weideman (2013a) voor dat akademiese diskoers as ’n diskrete materiële taalsfeer (“material lingual sphere”) benader moet word. Ooreenkomstig hierdie siening word taal nie alleen deur die spesifieke konteks of situasie waarin dit gebruik word (d.i. die feitelike situasie), bepaal nie, maar ook deur normatiewe beginsels wat logies, esteties, sosiaal, eties, juridies, ekonomies, tegnies of konfessioneel van aard is. Diskoers wat as ’n materiële taalsfeer benader word, kan gedefinieer word as “a system of typical lingual norms that regulate typical lingual facts [...] within the defining and limiting context of a socially differentiated lingual sphere” (Patterson & Weideman, 2013a:110).

Ten einde ’n spesifieke materiële taalsfeer te identifiseer, moet daar vasgestel word of die taalgebruik in ’n bepaalde konteks op unieke wyse van die taalgebruik van ander sfere of diskoerstipes verskil, wat gekwalifiseer word deur byvoorbeeld die logiese, estetiese, juridiese, etiese, konfessionele, ekonomiese, sosiale of ander aspekte of modaliteite van die werklikheid. So byvoorbeeld word die taalgebruik in ’n preek nie slegs deur die spesifieke situasie waarin dit plaasvind (bv. ’n kerkdienst) gekwalifiseer nie, maar ook deur die konfessionele aspek van die werklikheid wat die normatiewe beginsel verskaf waarop die konkrete taalgebruik in sodanige konteks sal reageer of antwoord. Hiervolgens lê die eiesoortigheid van ’n bepaalde diskoerstipe dus opgesluit in die spesifieke ervaringsdimensie waardeur dit gerig of gelei word.

In die geval van akademiese diskoers is die kwalifiserende dimensie wat hierdie taalgebruik normatief bepaal en van ander diskoerstipes onderskei die analitiese modaliteit van die werklikheid. Patterson en Weideman (2013a:111) ekspliseer hierdie idee soos volg:

While each academic field is circumscribed by one or more modes of reality (for example, mathematics is related to an investigation of the numerical and spatial modes, psychology is related to the psychological, or sensitive, mode), academic discourse as a whole is qualified by the analytical (or logical) mode, which is usually historically grounded.

Met ander woorde, hoewel die verskeie dissiplines binne akademiese diskoers ook deur 'n logies-analitiese bril na ander aspekte van die werklikheid kyk (d.i. numeries, esteties, ekonomies, ens.) as invalshoeke vir teoretiese begripsvorming, is die logiese dimensie, met analise as bepalende kern, die enkel oorkoepelende aspek en wesenstrek van akademiese taalgebruik in die geheel. Akademiese diskoers se eiesoortige karakter is afkomstig van die feit dat dit in die eerste plek analities of logies van aard is, en tweedens, dat dit 'n tipe diskoers is wat put uit vormende invloede van dít wat dit voorafgegaan het, d.i. elke akademiese dissipline is histories of tegnies gefundeer. Die wisselwerking tussen hierdie twee funksies kan skematies soos volg voorgestel word (oorgeneem uit Patterson & Weideman, 2013a:112):



*Figuur 3.1 Terminale funksies van akademiese diskoers*

Die gevolgtrekking waartoe hierdie outeurs op grond van die bostaande kom, is dat die eiesoortigheid van akademiese diskoers in die maak van onderskeidings gesetel is, 'n aktiwiteit wat primêr met die analitiese ervaringsdimensie in verband gebring kan word. Hoewel hulle hierdie gevolgtrekking slegs as voorlopig beskryf, met dien verstande dat dit in die toekoms aan verdere

bestudering en uitbouing onderwerp sal moet word, verskaf dit nietemin 'n beter begrip van die eiesoortigheid van akademiese taalgebruik wat, soos reeds vermeld, as deurslaggewend vir die hersiening van 'n konstruk van akademiese geletterdheid beskou word. Die vraag wat gevolglik ontstaan, is op welke manier dit sodanige konstruk kan verryk of toelig? Patterson en Weideman (2013b:137) voer in hierdie verband aan dat “[i]f the typicality of academic discourse is derived from the distinction-making activity associated with the analytical or logical mode of experience, then that aspect needs to take precedence”.

Hoewel hierdie aspek miskien nog nie prominent genoeg in die oorspronklike weergawe van die toetskonstruk figureer nie, is dit geensins op die oomblik totaal afwesig nie. Patterson en Weideman (2013a:118-119) lig uit hoedat verskeie van die huidige konstruk se komponente trouens reeds eksplisiet of implisiet na die vermoë verwys om op logiese en analitiese wyse met taal om te gaan. So byvoorbeeld word daar melding gemaak daarvan om onderskeid te tref tussen noodsaaklike en nienoodsaaklike inligting, feite en menings, stellings en argumente, oorsaak en gevolg, asook van analitiese vermoëns soos vergelyking, klassifisering en kategorisering. So ook maak die vermoë om opeenvolging en volgorde raak te sien, staat op logiese analise of onderskeiding. Waar daar in die konstruk na die argumentatiewe aard van akademiese taal verwys word, skemer die invloed van die analitiese modaliteit ook kenbaar deur. Aspekte soos samehang en logiese ontwikkeling, asook bewysvoering, afleiding en ekstrapolering word byvoorbeeld vermeld. Hieruit blyk duidelik dat akademiese argumentering is “neither of a juridical, nor of a political, promotional, ethical or confessional kind, but is of service [...] to the overriding analytical ends to which it is put” (Patterson & Weideman, 2013a:119). Ook die kommunikatiewe funksies wat in die konstruk uitgelig word, ondersteun die analitiese karakter van akademiese diskoers. Patterson en Weideman (2013a:119) wys egter daarop dat

[w]hile distinction-making and analytical or logical thinking are part of this particular definition, there is a possibility that they are not yet sufficiently foregrounded as the most important aspects of academic literacy. If constructs depend on ideas [...] then this foregrounding should be extremely important.

Na hulle mening is ten minste drie tipes wysigings aan die oorspronklike konstruksie toepaslik en geskik: eerstens, groter klem op die analitiese wesen van akademiese taalgebruik; tweedens, 'n duideliker artikulasie van komponente wat tans slegs geïmpliceer word, en derdens, 'n beter houding op die aard van akademiese interaksie deur middel van taal, wat die inwin, prosessering en skep van inligting kan insluit. Patterson en Weideman (2013b:139-140) stel vervolgens die volgende verfynde weergawe van die oorspronklike konstruksie voor, wat as basis vir die ontwerp van TAGNaS dien:

Akademiese geletterdheid kan beskryf word as die vermoë om:

- krities te dink (analiseer en oorweeg die gebruik van tegnieke en argumente) en logies en sistematies te redeneer in terme van jou eie en ander se navorsing;
- onderskeidings te maak tussen noodsaaklike en nienoodsaaklike inligting, feite en menings, stellings en argumente en om onderskeid te tref tussen oorsaak en gevolg, om data wat vergelykings tref te klassifiseer, kategoriseer en hanteer;
- met (gesproke en geskrewe) tekste in wisselwerking te tree: bespreek, bevraagteken, stem saam/verskil, evalueer, ondersoek probleme, analiseer, bring tekste met mekaar in verband, maak logiese gevolgtrekkings uit tekste, en skep dan nuwe tekste;
- inligting uit veelvuldige bronne met jou eie kennis te sintetiseer en te integreer ten einde nuwe idees te skep, met begrip van akademiese integriteit en inagneming van die risiko's van plagiaat;
- verbande tussen verskillende dele van 'n teks te verstaan, bewus te wees van die logiese ontwikkeling en organisasie van 'n (akademiese) teks vanaf inleiding tot gevolgtrekking, en te weet hoe om taal te gebruik wat die verskillende dele van 'n teks laat saamhang;
- te weet wat as bewyse vir 'n argument geld, om uit inligting te ekstrapoleer deur afleidings te maak en die inligting of die implikasies daarvan op ander gevalle as die een voorhande toe te pas;

- kreatief te dink: verbeeldingryke en oorspronklike oplossings, metodes of idees wat die gebruik van 'n dinkskrum of breinkaarte, visualisering en assosiasie kan behels;
- inligting in grafiese of visuele formaat te interpreteer, te gebruik en te skep;
- 'n reeks akademiese woordeskat te verstaan en te gebruik, asook inhoud of dissipline-spesifieke woordeskat in konteks;
- metafore en idioome vir akademiese gebruik te interpreteer, asook om konnotasie, woordspeling en dubbelsinnigheid raak te sien;
- verskillende tekstipes (genres) te interpreteer en begrip te toon vir die betekenis wat daardeur oorgedra word en die gehoor op wie dit gemik is;
- gespesialiseerde of komplekse grammatikale strukture te kan gebruik; 'n groot leksikale verskeidenheid te beheers, asook formele, gesaghebbende uitdrukkings, en abstrakte/tegniese konsepte te kan gebruik;
- betekenis te skep (bv. van 'n akademiese teks) verby sinsvlak;
- opeenvolging en volgorde raak te sien, om eenvoudige numeriese skattings en berekeninge te doen wat op akademiese inligting betrekking het, die tref van vergelykings toelaat en vir argumentasiedoeleindes gebruik kan word;
- jou lees en skryf te interpreteer en aan te pas vir analitiese/argumentatiewe doeleindes en/of in die lig van jou eie ervaring;
- die kommunikatiewe funksie van verskeie uitdrukkingswyses in akademiese taal te verstaan (soos definiëring, die verskaffing van voorbeelde, die maak van afleidings, ekstrapolering, argumentering); en
- op 'n gesaghebbende manier te skryf, wat behels die teenwoordigheid van 'n "ek" wat 'n denkbeeldige gehoor van spesialiste/nuwelinge (jong navorsers)/die algemene publiek/media aanspreek.

Patterson en Weideman (2013b:140) benadruk dat die verwerkte weergawe hierbo bloot die resultaat van 'n klemverskuiwing is; dit kom eerder op 'n verfyning neer as wat daar 'n heel nuwe koers met hierdie definisie ingeslaan

word. Hulle meen nietemin dat “though it may not constitute a major departure from the existing definition of academic literacy [...] [it] provides a more adequate overview of the requirements of academic literacy” (Patterson & Weideman, 2013b:140). Op die koop toe daag dit toetsontwerpers ook uit om aanpassings, selfs al is dit in geringe mate, met betrekking tot die operasionalisering van die konstruk te maak. Patterson en Weideman (2013b:140-145) maak in dié opsig verskeie voorstelle vir wysigings aan bestaande subtoetse en taaktipes van TALL, TAG en TALPS, asook vir die skep van en eksperimentering met nuwe soorte. Dit sal weldra uit die daaropvolgende paragrawe duidelik word watter vernuwende elemente in die loodstoetsontwerp van TAGNaS geïnkorporeer word, meer spesifiek uit die afdeling wat ’n oorsig van die verskillende subtoetse in TAGNaS gee. Die fokus van die bespreking verskuif dus vervolgens van die artikulasie van die betrokke toetskonstruk na die operasionalisering daarvan.

### **3. TAGNaS as ’n potensiële driefasetoets**

Soos reeds in Hoofstuk 1 aangedui is, onderneem hierdie studie om ’n eerste tree te gee tot die moontlike implementering van ’n meerfasige (in plaas van eenfasige) akademiese geletterdheidstoets, spesifiek in die vorm van ’n driefasemodel (“three-tier test model”). Met ’n driefasetoets word hiér bedoel ’n toets waarvan die volledige afhandeling in drie stadia of fases verloop, m.a.w. ’n toets wat in sy geheel nie in een sitting afgelê word (soos wat tans met bv. TALPS die geval is) nie, maar in drie afsonderlike sittings, hoewel elke fase na regte as ’n toets in eie reg beskou sou kon word. ’n Driefasemodel lyk gewis op die oog af moeiliker om te implementeer en te administreer, maar skep inderwaarheid ook die geleentheid vir verskeie alternatiewe operasionaliseringsmoontlikhede wat andersins nie haalbaar sou wees nie, en potensieel selfs tot verhoogde effektiwiteit en bruikbaarheid (’n duidelik ekonomiese analogie in die tegniese sfeer) aanleiding kan gee. Die ontstaan van sulke moontlikhede kan hoofsaaklik toegeskryf word daaraan dat die vereistes

van en uitslag vir 'n vorige toets in 'n driefasemodel 'n effek op die aflegging van 'n volgende een in die reeks kan hê, 'n wederkerigheid waarin allerlei toetsingsopsies opgesluit lê. So byvoorbeeld bied 'n driefasemodel die moontlikheid van verpligte/opsionele siftingstoetse wat gebruik kan word om te bepaal of dit hoegenaamd nodig is dat kandidate die daaropvolgende toetsfase(s) moet voltooi al dan nie – 'n tipe prosedure wat myns insiens eerder moeite- en tydbesparend as meer omslagtig is.

Die operasionalisering van 'n driefasemodel wat voorlopig vir TAGNaS in die vooruitsig gestel word, sien soos volg daaruit:

	<b>Getal subtoetse</b>	<b>Toetsmetode</b>	<b>Puntetotaal</b>	<b>Voorgestelde tyd</b>	<b>Opsioneel/Verpligtend/Vrygestel/Grensgevalle</b>
<b>Fase 1</b>	3	Veelkeusevrae	40	30 min.	Opsioneel/Verpligtend
<b>Fase 2</b>	9	Veelkeusevrae	100	90 min.	Verpligtend/Vrygestel
<b>Fase 3</b>	1	Skryftaak (Opstel)	60	60 min.	Vrygestel/Grensgevalle

*Tabel 3.3 TAGNaS as 'n potensiële driefasetoets*

Die motivering vir die eksperimentering met 'n driefasemodel in die ontwerp van TAGNaS kom hoofsaaklik vanuit drie oorde. In haar meestersverhandeling oor nagraadse akademiese geletterdheidstoetsing voer Anna Pot (2013) 'n diagnostiese analise van TALPS-resultate uit met die doel om aan te dui hoe sodanige ontleding toetsontwerpers nader kan toelig omtrent die probleme wat studente met betrekking tot akademiese geletterdheid en akademiese skryfwerk ervaar. Sy bevind onder andere dat die meeste van die opstelle wat in antwoord op die akademiese skryfvraag in TALPS geproduseer word, onvoltooid of onafgerond is. Hoewel tyd ook 'n indikator van vaardigheid in akademiese diskoers is, is sy van mening dat “[time] is now an important if not the main deciding factor through which many essays fall short in their ability to display an adequate awareness of what constitutes academic literacy” (Pot, 2013:54). Om hierdie struikelblok uit die weg te ruim, beveel sy onder andere aan dat die idee van 'n meerfasige toetsmodel in toekomstige weergawes van TALPS ontgin



behoort te word. Vir TALPS – in sy huidige formaat – stel sy spesifiek ’n tweefasemodel voor, met ’n eerste toets wat die sewe veelkeuseafdelings bevat en ’n tweede toets wat uitsluitlik uit die akademiese opstelvraag bestaan. Sy maak die verdere voorstel dat alle kandidate verplig moet word om die eerste toets te skryf, en diegene wat benede die huidige afsnypunt van 60% (minder risiko) behaal, die geleentheid gegun moet word om die opsteltoets wat as ’n tweede kans kan dien, af te lê – ’n tipe strategie wat ook vir TAGNaS ontplooi kan word. Vir TAGNaS word die opsie van ’n addisionele, korter siftingstoets egter ook oorweeg, ’n toets waarvan die afsnypunt nóg hoër as 60% gestel kan word (bv. min tot geen risiko) ten einde die sterk, bedrewe studente vroeër uit die toetsingsproses uit te skakel.

Volgens Pot (2013:55-56) hou ’n meerfasige model soos dié verskeie voordele in. Omdat die opsteltoets in so ’n model op ’n ander tydstip as die veelkeusetoets(e) afgelê word, laat dit studente toe om ten volle op die skryftaak voorhande te konsentreer, sonder om uitgeput te voel ná die beantwoording van ’n hele aantal direk voorafgaande veelkeusevrae. Verder kan die vereiste lengte van die opstel wat geskryf moet word van die huidige 300-400 na 500-800 woorde verhoog word. Pot (2013:55) se diagnostiese analise dui aan dat studente wat wel die opstelvraag volledig voltooi, dit soms moeilik vind om ’n goed gestruktureerde, uitgebreide argument in slegs 300 woorde uit te druk. Boonop kan ’n langer opstel ’n akkurater aanduiding van ’n student se akademiese skryfvermoë gee wat weer tot meer akkurate assessering kan lei, ’n aspek wat ’n positiewe invloed op die betroubaarheid en effektiwiteit van ’n toets kan hê. In ’n aparte opsteltoets kan meer bronmateriaal ook aan studente voorsien word (bv. een ekstra teks oor die tema), bo en behalwe die tekste waaraan studente reeds in die voorafgaande veelkeusetoets(e) blootgestel is. Laastens kan ’n tweede (of derde) toets ook as ’n tweede (of derde) kans dien vir kandidate wat moontlik negatief deur ’n voorafgaande toets geaffekteer is deurdat hulle foutiewelik as potensiële risikostudente geklassifiseer word. Pot (2013:56) wys

daarop dat foutiewe klassifikasie in enige toets kan gebeur, aangesien geen toets ooit as 'n volledig betroubare meetinstrument bestempel kan word nie.

Die tweede stimulus vir proefneming met 'n meerfasige toetsontwerp in hierdie studie is die bewese sukses van die tweefasige Diagnostic English Language Needs Assessment (DELNA) wat by die Universiteit van Auckland in Nieu-Seeland geïmplementeer is. Hierdie taaltoets bestaan uit twee toetse waarvan die eerste 'n 30 minuut-siftingstoets en die tweede 'n twee uur lange diagnostiese toets is. Die siftingstoets moet deur alle kandidate afgelê word, is rekenaargebaseerd en bestaan uit woordeskat oefeninge en enkele tydsgebonde leestake. Op grond van die uitslag wat behaal word, kan kandidate die terugvoering kry dat hulle akademiese taalvaardighede geskik vir universitêre studie is; dat hulle akademiese taalvermoë bevredigend is, maar wel opknapping nodig het, óf dat hulle die tweede, diagnostiese taaltoets óók sal moet aflê sodat 'n meer gedetailleerde profiel van hulle taalvaardigheidvlakke verkry kan word. Teenoor die siftingstoets word die diagnostiese toets beskryf as 'n deeglike pen-en-papier-assessering wat 'n lees-, luister- en skryfaktiwiteit bevat (DELNA, 2016).

Ten laaste bied Geldenhuys (2007) 'n derde motivering vir die oorweging van 'n alternatief vir die konvensionele enkeltoetsformaat. Hy herroep hoedat daar destyds by die ontstaan van TALPS reeds met die idee van twee afsonderlike toetse rondgespeel is, waarvan die eerste, net soos by DELNA, 'n siftingsfunksie sou verrig. Volgens Geldenhuys (2007:74) was die aanvanklike betoog vir 'n korter toets in die geval van TALPS die volgende:

Given the potentially heterogeneous test population – students that register for postgraduate studies [...] vary from those who are very literate in English to students who come from, for example, French-speaking African countries and countries in Asia and who are not that literate in English – we needed an instrument to give a first, rough indication of level in order not to waste the time of those competent enough to make a success of their studies. The line of thought therefore was to administer a shorter test to all candidates, and after receiving the results of this test administer the longer test to candidates who were potentially at risk of not completing their studies in the desired time.

Geldenhuis (2007:74) besin vervolgens oor die lewensvatbaarheid van so 'n korter siftingstoets. 'n Langer toets is in die reël potensieel meer betroubaar as 'n korter toets. Aan die ander kant weer boet toetse wat beide betroubaar en geldig (en in hierdie sin dus effektief) is, maar te lank is (bv. vir die doeleindes van vinnige sifting) in aan bruikbaarheid. Daar is hier dus 'n wisselspel van die analogies kinematiese, fisiese en ekonomiese momente in die tegniese gekwalifiseerde metingsproses, waarin daar tegniese afweging en beoordeling gebruik moet word om respektiewelik die verlangde mate van betroubaarheid, die graad van geldigheid, en die vlak van nuttigheid op te weeg. In die geval van korter siftingstoetse sal ontwikkelaars dus hoogs waarskynlik met laer (maar steeds aanvaarbare) betroubaarheidswaardes tevrede moet wees. Soos reeds in Hoofstuk 2 opgemerk is, vereis verantwoordelike toetsontwerp baiekeer dat redelike toegewings tussen (dikwels botsende) ontwerpvereistes gemaak moet word. Sulke tegniese kompromieë word begripsmatig beter verken deur te verreken dat daar 'n analogiese skakel is tussen die tegniese en die ekonomiese dimensies van ons ervaring.

Uit daardie analogiese skakel spruit die beginsel van tegniese afweging en bruikbaarheid. Dieselfde beginsel is van toepassing wanneer keuses gemaak moet word omtrent watter subtoetse of komponente in die korter toets ingesluit, en watter daaruit weggelaat, behoort te word. Om suiwer op grond van empiriese data te besluit, is volgens Geldenhuis (2007:81), nie noodwendig die verstandigste roete om te volg nie, “even this kind of decision about what should make up a shorter test is, in the end, one that must be done judiciously, i.e. after weighing up various issues”. Daar is dus 'n verdere tegniese analogiese moment ter sprake, voortspruitend uit die verband tussen die tegniese en die juridiese: die beginsel is dat die oorweging van wat gekies word en wat agtergelaat word, steeds reg moet laat geskied aan die doel van die toets en die vermoë wat gemeet word.

Wat die beoogde driefaseontwerp van TAGNaS betref (kyk tabel 3.3 en 3.4), is dit nodig om, in die lig van die bostaande bespreking, 'n paar belangrike opmerkings te maak. Wat voorlopig vir TAGNaS in die vooruitsig gestel word, is dat die eerste 30 minuut-veelkeusetoets, net soos in die geval van DELNA, as 'n siftingstoets (hetsy verpligtend of opsioneel) ingespan word, ook moontlik in elektroniese formaat. Die drie subtoetse waarop tussentyds vir die eerste loodswaerking van hierdie toets besluit is, sluit in skommeltekste, woordeskatitems en 'n grammatika en teksverbandoefening – raadpleeg die toetsspesifikasies in die volgende afdeling vir 'n verduideliking van elk van hierdie subtoetse. Hierdie spesifieke keuses, hoewel nog slegs preliminêr en onderhewig aan voldoende empiriese proefneming, is nie om dowe neutemaak nie. Allerlei oorwegings, van teoretiese en praktiese aard, is in berekening gebring. Omdat die toets uit slegs veertig punte tel en binne dertig minute voltooi moet kan word, noodsaak dit die insluiting van korter tipe subtoetse wat binne die toegelate tyd afgehandel behoort te kan word. Veral die woordeskatitems, 'n tipe subtoets wat ook in die DELNA-siftingstoets gebruik word, is nuttig omdat dit kort items is wat nie noodwendig indringende bestudering en aandag, soos in die geval van teksbegrip (wat gewoonlik die langste subtoets is) en die interpretasie van visuele inligting, vereis nie. Daarbenewens is Du Plessis (2012:56) van mening dat “the assessment of vocabulary is one of the main specifications of a test of academic literacy”. Verder is die grammatika en teksverband-subtoets, wat ook van TALL, TAG en TALPS se ontwerpe deel uitmaak, gewoonlik 'n baie betroubare subtoets wat tradisioneel goed tussen kandidate diskrimineer (Van Dyk & Weideman, 2004b). Soos uit die toetsspesifikasies behoort te blyk (kyk bylaag B), dek hierdie drie subtoetse saam ook verskeie van die komponente van die toetskonstruk wat in die vorige afdeling bespreek is.

Verdere ondersteuning in die vorm van empiriese data en analise is egter nodig om met groter sekerheid te kan sê in watter mate hierdie (voorlopige) seleksie van subtoetse (vir siftingsdoeleindes) teoreties geregverdig is al dan nie.

Sodanige data kom wel in die volgende hoofstuk van hierdie studie aan bod, maar daar moet in gedagte gehou word dat hierdie studie slegs 'n eerste rondte van eksperimentering met 'n driefasetoetsformaat verteenwoordig. Hoewel daar stellig aan 'n voldoende aantal toetsitems oorsprong gegee kon word om eksperimentering met 'n driefaseontwerp moontlik te maak, sal die empiriese bewyse wat binne die bestek van hierdie studie gegenerer kon word deur verdere navorsing aangevul móét word. Geldenhuys (2007) stel bv. die uitvoering van verskeie korrelasiestudies voor. Soos wat verdere data beskikbaar kom, kan prestasie in die siftingstoets met prestasie in die daaropvolgende tweede en/of derde toetse gekorreleer word ten einde die voorspellingsgeldigheid daarvan beter te bepaal of te bevestig.

Ten slotte volg 'n meer gedetailleerde uiteensetting van die driefaseontwerp wat vir TAGNaS beoog word, mét 'n aanduiding van die tipe subtoetse wat vir elke toetsfase beplan word. Die verskillende subtoetse word in meer besonderhede in die volgende afdelings bespreek.

	Subtoetse	Punte-toekenning	Totaal	Tyd	Toets-metode	Opsioneel/Verplichtend/Vrygestel/Grensgevalle
<b>Fase 1</b>	Skommeltekste	10	40	30 min.	Veelkeuse	Opsioneel/Verplichtend
	Akademiese woordeskat	12				
	Grammatika en teksverband	18				
<b>Fase 2</b>	Skommelteks	5	100	90 min.	Veelkeuse	Verplichtend/Vrygestel
	Interpretasie van visuele inligting	12				
	Akademiese woordeskat	12				
	Tekstipe	5				
	Teksbegrip	30				
	Vergelyking tussen tekste	10				
	Grammatika en teksverband	16				
	Opsomming	5				
	Verwysings	5				
<b>Fase 3</b>	Argumenterende opstel	60	60	60 min.	Skryftaak	Vrygestel/Grensgevalle

Tabel 3.4 TAGNaS as 'n potensiële driefasetoets (met subtoetse)

#### 4. Toetsspesifikasies van TAGNaS

Ná die keuse van 'n toetskonstruk is die volgende belangrike stap in die ontwerpproses die belyning van die konstruk met toetsspesifikasies en die gepaardgaande toetsitems. Toetsspesifikasies sit uiteen wat 'n toets ontwerp is om te meet en hoe dit gemeet gaan word. Hoewel dit verskeie formate kan aanneem, bevat toetsspesifikasies gewoonlik inligting oor die doel van die toets, die teikengroep, die toetsinhoud en -formaat, asook ander relevante besonderhede soos voorgestelde tydsduur, puntetoekenning, ens. (Davies et al., 1999:207). Daar word dikwels na toetsspesifikasies as 'n bloudruk vir toets- of itemontwikkelaars verwys. Bachman en Palmer (2010:368) beskryf dit as die skakel tussen die ontwerpstadium en die assessering self, sowel as die primêre dokument wat aan toetsontwikkelaars leiding gee in die skep van assesseringstake. Dit help om te verseker dat “the test developer’s intentions [...] are actually implemented in the creation of assessments” (Bachman & Palmer, 2010:366).

'n Bloudruk speel egter nie net in die produksie van 'n meetinstrument 'n belangrike rol nie, maar ook in die verdediging daarvan. Davies et al. (1999:207) dui aan dat toetsspesifikasies deurslaggewend vir die opbou van 'n toets se konstrukgeldigheid is, waarby inhoudsgeldigheid ook ondergebring kan word. Inhoudsgeldigheid word bereik wanneer die toetsinhoud die spesifieke domein of konstruk wat gemeet word, bevredigend verteenwoordig, d.i. dat al die vernaamste aspekte wel in geskikte proporsies (ook in die bloudruk) gedek word. Toetsspesifikasies bevorder verantwoordelike toetsontwerp deurdat dit, soos Bachman en Palmer (2010:369) dit stel, 'n basis vir gehaltebeheer gedurende die produksie van toetsitems verskaf:

[B]y continuously monitoring the creation of different assessment tasks and different forms of the assessment, the test developer and the test development team can correct any discrepancies they find between the specifications in the Blueprint and the characteristics of the assessment tasks and different forms of the assessment.

In terme van die teoretiese raamwerk waarop hierdie studie berus, is die saamstel van 'n bloudruk 'n regulatiewe ontwerpvereiste wat voortspruit uit die koppeling van die leidende tegniese funksie met die linguale dimensie van die werklikheid: die tegniese ontwerp moet dus uitgedruk ('beteken') of geartikuleer word in die formaat van 'n bloudruk. Daarbenewens werk toetsspesifikasies ook mee tot die verwesenliking van die estetiese ontwerpmoment wat onder andere inhou dat belyningsfoute (of diskrepansies) tussen die verskillende komponente van 'n toets sover moontlik verhoed of reggestel behoort te word. Sulke tegniese harmonisering sal bv. verhoed dat verskillende en weersprekende perspektiewe op wat taalvermoë behels, soos die beperkende en verruimde visies op taal wat in Tabel 3.1 hierbo vermeld is, lukraak en deurmekaar gebruik word as basis vir die toetskonstruk en die -bloudruk wat daaruit voortspruit.

TALPS se formaat (wat op sy beurt weer met dié van TALL korrespondeer) lê die grondslag vir die toetsspesifikasies van TAGNaS, hoewel vernuwend elemente ook figureer vanweë die hersiene konstruk waarop die ontwerp gebaseer word – meer hieroor in die volgende afdeling. Vir 'n uiteensetting van die toetsspesifikasies van beide TALPS en TAGNaS raadpleeg Bylaag A en B onderskeidelik.

Die onderstaande statistiek, soos voorberei deur Iteman 4.2, is 'n duidelike bewys daarvan dat die bloudruk wat vir TALPS geskep is oor die algemeen effektief is en, spesifiek in die geval van hié die studie, beslís 'n navolgenswaardige prototipe verskaf.

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0.932	3.583	0.873	0.659	0.884	0.932	0.795	0.938
Scrambled text	0.829	0.772	0.503	0.680	0.674	0.669	0.810	0.805
Graphic and visual	0.831	1.231	0.668	0.669	0.705	0.801	0.801	0.827
Vocabulary	0.549	1.288	0.364	0.352	0.353	0.533	0.521	0.522
Text types	0.668	0.897	0.595	0.614	0.294	0.746	0.761	0.454
Understanding texts	0.788	1.811	0.601	0.606	0.683	0.751	0.755	0.812
Grammar and text relations	0.882	1.476	0.743	0.665	0.798	0.852	0.799	0.888
Text editing	0.902	0.917	0.772	0.732	0.764	0.872	0.845	0.866

*Tabel 3.5 Betroubaarheidswaardes van die subtoetse in TALPS soos afgeleë in 2011 en 2012 by die Universiteit van die Vrystaat (UFS)*

Die gekombineerde Alpha-waarde van 0.932 hierbo getuig van die hoë kwaliteit van TALPS en sy onderskeie subtoetse. Volgens algemeen aanvaarde toetspraktyk kan 'n betroubaarheidswaarde van 0.7 of hoër as voldoende geag word. Soos uit tabel 3.6 sigbaar is, lewer die verskillende subtoetse in TALPS ook hoogs bevredigende fasiliteits- (Mean P) en diskriminasiewaardes (Mean Rpbis) op. Goeie toetsitems is daartoe in staat om tussen sterk en swak kandidate te onderskei en sal daarom 'n hoër Rpbis-waarde (wat van .0 tot 1.0 kan wissel) hê, maar selde hoër as 0.50 (Guyer & Thompson, 2011). Die gekombineerde Rpbis-waarde van 0.328 stel dus in 'n hoë mate tevrede, so ook die gemiddelde P-waarde van 0.628 wat binne die gestelde parameters van 0.15 (min.) en 0.84 (maks.) val.

Score	Items	Mean	SD	Min Score	Max Score	Mean P	Mean Rpbis
All items	76	47.737	13.744	2	76	0.628	0.328
Scored Items	76	47.737	13.744	2	76	0.628	0.328
Scrambled text	5	2.354	1.868	0	5	0.471	0.269
Graphic and visual	10	5.754	2.992	0	10	0.575	0.406
Vocabulary	10	7.146	1.918	0	10	0.715	0.231
Text types	5	2.442	1.556	0	5	0.488	0.230
Understanding texts	21	14.514	3.933	0	21	0.691	0.289
Grammar and text relations	15	8.371	4.292	0	15	0.558	0.414
Text editing	10	7.157	2.924	0	10	0.716	0.380

*Tabel 3.6 Beskrywende statistiek vir die subtoetse in TALPS soos afgeleë in 2011 en 2012 by die Universiteit van die Vrystaat (UFS)*



In die volgende afdeling van hierdie hoofstuk word 'n oorsig gegee van die verskillende subtoetse wat ooreenkomstig die toetsspesifikasies in Bylaag B in die eerste loodswaerking van TAGNaS ingesluit is, asook van die aspekte van akademiese geletterdheid wat deur elk getoets word. Daar word pertinent uitgelig hoe die subtoetse, waarvan sommige ook gewysigde of nuwe subtoetse is, met die analitiese karakter van akademiese diskoers, wat aan die kern van die toetskonstruk lê, bely is. 'n Gulde reël wat aangaande die keuse van subtoetse in gedagte gehou moet word, is dat hoe groter die verskeidenheid van subtoetse ('n biologiese analogie binne die tegniese sfeer), hoe groter is die kans gewoonlik dat die meetinstrument betroubaar sal toets (Van Dyk & Weideman, 2004b).

## **5. Die subtoetse van TAGNaS**

Die subtoetse wat vir die verskillende fases van TAGNaS ontwerp en daarna aan loodstoetsing onderwerp is, sluit die volgende in:

### *1. Skommeltekste*

Hierdie subtoets vereis van kandidate om die korrekte volgorde te bepaal van 'n reeks sinne in 'n paragraaf wat geskommel/deurmekaar aangebied word. Die uitvoering van hierdie taak berus kennelik op die toepassing van logiese analise, soos deur Patterson en Weideman (2013b:140) beaam dat “[it] is clearly a measure of [...] the ability to distinguish not only different and successive analytical points (evidence) made in a text, but also their systematic progression”. Butler (2009:294) meen verder dat hierdie subtoets “not only [tests] students’ ability in recognising text relations, drawing on their interpretative abilities regarding the context, but also their ability to recognise lexical clues contained in the sentences”.

### *2. Interpretasie van visuele inligting*

In hierdie subtoets beantwoord kandidate vrae wat op die interpretasie van 'n grafiek gebaseer word en voer ook eenvoudige numeriese bewerkings uit.

Analitiese denke tree hier sterk na vore deurdat vaardighede soos onderskeiding, vergelyking, afleiding (bv. die identifisering van tendense en proporsies) en bewyslewing betrek word. Hierdie subtoets, asook die skommelteks-subtoets hierbo, maak beide van ook onder andere TALPS se ontwerp deel uit.

### *3. Akademiese woordeskat*

Hierdie subtoets toets kandidate se kennis van algemene akademiese woordeskat aan die hand van veelkeuseitems wat op grond van 'n akademiese woordeskatlys ontwerp word. Vir die woordeskatitems in die Engelse akademiese geletterdheidstoets (soos TALL en TALPS) maak die ontwikkelaars van die Inter-Institutional Centre for Language Development and Assessment (ICELDA) vir 'n geruime tyd lank reeds van die Academic Word List (AWL) van Coxhead (2000) gebruik. Vir die Afrikaanse toets (bv. TAG) is daar vanweë die gebrek aan 'n soortgelyke lys in Afrikaans tot onlangs toe nog, benewens Coxhead (2000), ook op Nederlandse bronne soos die *Woordenlijst Academisch Nederlands* en *Wijze woorden* van Giezenaar en Schouten (2002) gesteun. Die benutting van 'n Nederlandse woordelys vir Afrikaanse toets kan egter, in die woorde van Steyn (2013), gewis meer vrae laat ontstaan as wat dit antwoorde verskaf. Daar is daarom vir die doeleindes van hierdie studie besluit om van Steyn (2013) se nuut ontwikkelde Akademiese Woordelys vir Afrikaans gebruik te maak wat redelik onlangs uit haar navorsing vir haar doktorsale proefskrif oor taaltoetsing voortgespruit het. Die woordeskat wat getoets word, omvat 'n seleksie van items wat uit die verskillende frekwensiekategorieë van Steyn (2013) se woordelys geneem is. Hierdie lys is deur Steyn (2013) ontwikkel in 'n poging om die leemte wat daar tans vir só 'n lys in Afrikaans bestaan, te vul. Meer volledige besonderhede oor die aard en samestelling van hierdie lys word later in hierdie hoofstuk verskaf.

Ofskoon Patterson en Weideman (2013b:140) die woordeskatsubtoets as 'n belangrike onderdeel van 'n akademiese geletterdheidstoets beskou, is hulle nogtans van mening dat hierdie taaktipe, soos wat dit tans in 'n toets soos TALPS

daar uitsien, “[...] might be more productively related to distinction-making”. In sy huidige formaat lyk ’n item wat van hierdie subtoets deel uitmaak, tipies soos volg:

Onlangse navorsing het aangetoon dat hoe meer \_\_\_\_\_ ’n ekostelsel is, hoe meer stabiel is die stelsel.

- A. uitgesluit
- B. uitvoerend
- C. uiteenlopend
- D. uitgesproke

Hoewel items gelykvormig aan die bogenoemde voorbeeld steeds in die loodstoets van TAGNaS opgeneem is, word daar ook, in antwoord op Patterson en Weideman (2013b) se kommentaar hierbo, met ’n gewysigde weergawe van dié taaktipe in hierdie studie geëksperimenteer. In die plek van ’n item wat slegs een ruimte vir ’n korrekte antwoord bevat, word daar in die aangepaste weergawe ’n dubbele gaping verskaf wat beide met korrekte opsies gevul moet word. ’n Voorbeeld van so ’n itemtipe is die volgende:

Daar is \_\_\_\_\_ onder marinebioloë dat daar \_\_\_\_\_ veranderings aan vissery-praktyke gemaak moet word om volhoubaarheid te verseker.

- A. konsensie ... ingewikkelde
- B. konsensus ... ingrypende
- C. konvensie ... integrale
- D. kontensieus ... interne

Die redenasie hier is dat hierdie wysiging, hoewel oënskynlik gering in omvang, heel waarskynlik tot ’n verskerpte aanspraak op kandidate se logiese onderskeidingsvermoë aanleiding sal gee. Nie net moet twee antwoorde, in plaas van een, korrek uit die konteks van die betrokke sin afgelei word nie, maar ook akkuraat uit die verskeie gelyste moontlikhede geïdentifiseer word. Of hierdie gewysigde taaktipe wel die potensiaal besit om produktiewe resultate op te lewer al dan nie, sal weldra uit die volgende hoofstuk in hierdie studie blyk. Om hierdie tipe woordeskatitem te skep, rig egter ’n groot uitdaging tot die ontwerpers daarvan. Die korrekte kontekstuele leidraad moet op so ’n wyse geformuleer

word dat dit van genoegsame betekenis is om die korrekte ontbrekende woordpaar daaruit af te kan lei, en terselfdertyd moet die afleiers (ander moontlike maar tog verkeerde antwoorde) 'n mate van geloofwaardigheid hê, sodat die korrekte keuse nie te vanselfsprekend en dus te maklik gaan wees nie.

#### *4. Tekstipe*

Hier moet kandidate twee stelle taaluitinge met mekaar vergelyk ten einde die uitinge wat tot dieselfde tekstipe of genre behoort met mekaar af te paar. Patterson en Weideman (2013b:140) wys daarop dat hierdie subtoets wat kandidate se bevatlikheid vir verskillende genres toets “already acknowledges the point of academic discourse being different in content from other discourse types”.

#### *5. Teksbegrip en 6. Vergelyking tussen tekste*

Die teksbegrip-subtoets assesser kandidate se begrip van 'n geskrewe teks (600-800 woorde) en is gewoonlik die langste subtoets in die hele versameling. Hierdie subtoets in sy huidige vorm lê reeds in groot mate klem op analitiese onderskeiding (afleiding en bewyslewing inkluis, vgl. die toetsspesifikasies in Bylaag B), maar kan ewenwel op verskeie maniere aangepas word om hierdie aspek meer op die voorgrond te laat tree. Een voorbeeld van 'n tipe item wat tipies van hierdie subtoets is, is die volgende:

Die hoofgedagte van paragraaf 1 is dat

- A. Hainan-gibbons se sang deesdae minder vrolik as gewoonlik is.
- B. die bewaring van woude op Hainan-eiland van geringe belang is.
- C. Hainan-eiland 'n ontwikkelende, maar kerngesonde ekonomie het.
- D. Hainan-gibbons weens menslike bedrywighede aan die uitsterf is.

Om hierdie subtoets egter nog nader met die verfynde toetskonstruk te belyn, word verskeie van Patterson en Weideman (2013b:140-142) se ontwerpaanbevelings in hierdie afdeling op die proef gestel. Byvoorbeeld, in ooreenstemming met die komponent van die toetskonstruk wat verwys na “met tekste in wisselwerking te tree: [...] analiseer, bring tekste met mekaar in verband

[...]” is daar uitdruklik toegesien dat items soos die volgende by hierdie afdeling van TAGNaS ingesluit word:

Die konklusie van die argument wat in paragraaf 4 die eerste keer genoem word, kan ons vind in die

- A. laaste sin van paragraaf 5.
- B. eerste sin van paragraaf 6.
- C. eerste sin van paragraaf 9.
- D. tweede sin van paragraaf 10.

Die gedagte dat bewaring en maatreëls spesifiek biodiversiteit kan bevorder, word in paragraaf 6 aan die orde gestel. Dit word verder geïllustreer in paragrawe

- A. 1, 2 en 3.
- B. 1, 3 en 5.
- C. 3, 5 en 10.
- D. 7, 8 en 9.

So ook word daar, deur die insluiting van ’n item soos die volgende, klem geplaas op die komponent van die konstruk wat heenwys na “die logiese ontwikkeling en organisasie van ’n (akademiese) teks”:

Die frase “die argument” in die laaste sin van paragraaf 4 wys na die uiteensetting daarvan in paragrawe:

- A. 1 tot 2.
- B. 2 tot 4.
- C. 3 tot 5.
- D. 5 tot 8.

In antwoord op die aanbeveling dat “[h]aving a range of texts that can be compared and contrasted [...] is indicated as desirable” (Patterson & Weideman, 2013b:142) word daar in TAGNaS met ’n bykomende teksbegrip-subtoets geëksperimenteer wat spesifiek vergelyking en/of kontras tussen tekste behels en boonop ook met die konstrukkomponent “inligting uit veelvuldige bronne [...] sintetiseer en [...] integreer” in verband gebring kan word. ’n Tweede, korter teks oor dieselfde onderwerp waarvan die standpunt aansienlik van die eerste, langer teks s’n verskil, word bykomend aan kandidate verskaf, met vrae soos die volgende wat daarop volg:

'n Kwessie wat deur die tweede teks aangeroeer word wat feitlik afwesig is in die eerste teks is

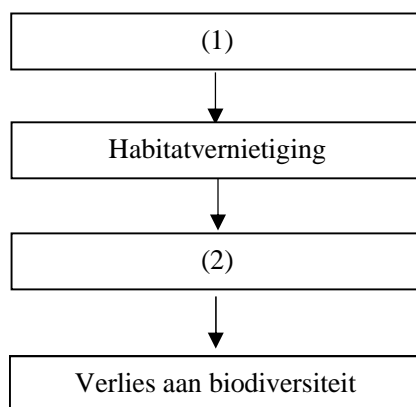
- A. 'n verhoging in lewenstandaard.
- B. die tempo van verlies aan diversiteit.
- C. die onafwendbaarheid van 'n ramp.
- D. 'n onverklaarbare sug na meer bronne.

Die eerste teks vermeld dat klimaatsverandering en die behoefte aan bewerkbare grond die grootste bydraers tot habitatvernietiging is. Watter bykomende faktor word in die tweede teks genoem?

- A. die mens se onversadigde vraag na grond
- B. die vergrotende ekologiese voetspoor van die mens
- C. Noord-Amerika se ekologiese voetspoor
- D. die uitputting van natuurlike hulpbronne

'n Verdere voorstel wat ook in 'n ander subtoets van TAGNaS (Opsomming) uitgetoets word, is die inkorporering van visuele analise, ook verwoord as die “visualisation of logical distinctions” (Patterson & Weideman, 2013b:143). 'n Loodsitem in die teksbegrip-subtoets wat bv. spesifiek op visuele onderskeidings berus, is die volgende:

Voltooi die diagram deur van die onderstaande inligting uit paragraaf 9 gebruik te maak.



Die woord of frase wat by (1) hoort, is:

- A. klimaatsverandering
- B. voedselvoorsiening
- C. uitsterwing van spesies
- D. oorlewing van spesies

Die woord of frase wat by (2) hoort, is:

- A. klimaatsverandering
- B. voedselvoorsiening
- C. uitsterwing van spesies
- D. oorlewing van spesies

Uit die bogenoemde wil dit lyk asof die teksbegripafdeling toelaat vir die inkorporering van heelparty van die bygewerkte komponente in die verfynde toetskonstruk. Daar word in die volgende hoofstuk oor die produktiwiteit al dan nie van die individuele items in die teksbegrip- en vergelykingsubtoetse, asook van hierdie subtoetse in hulle geheel, verslag gedoen.

### *7. Grammatika en teksverband*

Hierdie subtoets assesser die beheersing van sintaktiese verbande en potensieel ook die komponente van kommunikatiewe funksie, woordeskat en samehang. Dit behels dat kandidate moet bepaal waar woorde, asook watter woorde, uit 'n teks weggelaat is wat min of meer sistematies gemutileer is met behulp van sinsluiting. Hierdie subtoets word reeds met groot sukses in toetse soos TALL, TAG en TALPS gebruik.

### *8. Opsomming*

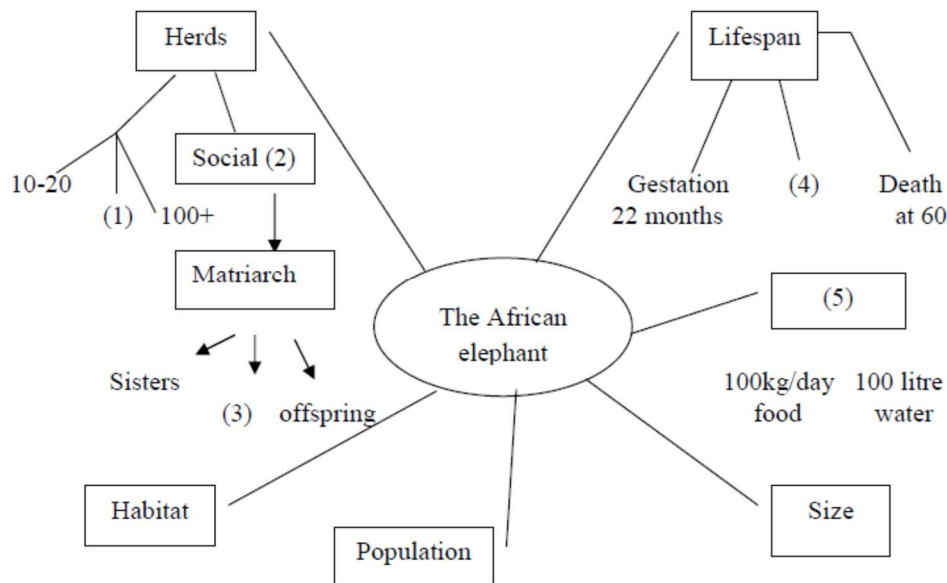
In hierdie subtoets voltooi kandidate 'n opsomming van 'n stuk teks in die vorm van 'n visuele diagram. Die fokus val dus hier op die visualisering van logiese onderskeidings, 'n subtoets waarvan die insluiting, volgens Patterson en Weideman (2013b:143), alreeds vantevore deur ontwikkelaars oorweeg, maar nog nie in die praktyk geïmplementeer of aan loodstoetsing onderwerp is nie. Vir hierdie studie is twee verskillende weergawes van hierdie subtoets vir die doeleindes van loodstoetsing geskep, beide geskoei op die prototipe van Patterson en Weideman (2013b) soos verskaf in tabel 3.7.

Listen to/read the following text, look at the diagram, and then answer the questions below:

**The African elephant**

Elephants essentially live in herds and may be found in groups of anything between 10 and 20 or up to 50 and more, and, in rare cases, in excess of 100. Their highly developed social structure, however, remains consistent throughout. Family units are led by a cow elephant, or matriarch, and a typical family herd consists of cow elephants of various ages: the leader, and her sisters, their daughters, and their offspring.

The lifespan of an elephant is long and often eventful. For one thing, elephants...



- (1) The most appropriate choice here is  
 A. between 30 and 40  
 B. more than 40  
 C. more than 50  
 D. about 70 or 80
- (2) The term that is used here is  
 A. development  
 B. structure  
 C. herd  
 D. family

Tabel 3.7 'n Voorbeeld van visuele analise (Patterson & Weideman, 2013b:143)

9. Verwysings

Hier word kandidate se kennis van verwysings getoets aan die hand van vrae gebaseer op 'n kort bronnelys wat voorsien word. Hierdie subtoets belyn nie alleen met 'n komponent soos “begrip van akademiese integriteit en die risiko’s



van plagiaat” in die konstruk nie, maar ook byvoorbeeld met ’n aspek soos “opeenvolging en volgorde”, spesifiek waar kandidate se vermoë om bronne in alfabetiese volgorde te lys, getoets word.

### *10. Akademiese skryf*

Hierdie subtoets bied aan kandidate die geleentheid om ’n geskrewe akademiese teks te produseer. Soortgelyk aan TALPS hou die tekste wat in TAGNaS gebruik word almal met dieselfde tema verband. Van kandidate word vereis om van enige van die inligting wat in die toets verskyn, gebruik te maak en ’n goed gestruktureerde argumentatiewe opstel van ongeveer 500-800 woorde oor die onderwerp wat verskaf word, te skryf. Omdat daar beoog word om die skryfopdrag as die derde toets van ’n driefasige model aan te bied, kan ’n addisionele teks oor die onderwerp, bo en behalwe die tekste wat in die vorige twee toetse verskyn, ook by hierdie subtoets ingesluit word. Kandidate moet toesien dat hulle by wyse van teksverwysings en ’n verwysingslys die nodige erkenning aan die bronne in die toets gee wat hulle kies om te gebruik. Hulle moet ook trou bly aan algemeen aanvaarde akademiese skryfkonvensies (soos formele register, logiese struktuur, bronverwysings, ens.). Hierdie subtoets vind veral nou aansluiting by die komponente van die toetskonstruk wat verwys na kritiese denke en onderskeidings, sistematiese redenering, interaksie met tekste, sintese, akademiese integriteit, teksorganisasie, argumentasie en eie akademiese stem. Pot (2013:20) toon aan hoedat akademiese skryf ’n komplekse taak is wat verskeie aspekte van akademiese geletterdheid, met analitiese onderskeiding as die kernkomponent, verenig:

The writing process therefore potentially involves many activities related to the ability to handle academic discourse. [...] [G]athering information is done either through reading or listening, making/reviewing notes or discussions with others; processing information is done through analysis, in which distinction-making (especially comparing, contrasting and categorising) plays an important part; and only then do we produce new information (finally, in writing) in which we state our opinion, that has been formed in the previous processes, and is the articulation of the result of our distinction-making or analytical activity. (Pot, 2013:20)

Vanweë praktiese beperkinge by die afneem van voorgraadse akademiese geletterdheidstoetse soos TALL en TAG word minder klem op die toets van

produktiewe akademiese skryfvermoë in hierdie toetse geplaas. Daar kan verder geargumenteer word dat by studente wat nuwe aankomelinge op universiteit is, akademiese skryfvermoë nog nie volledig ontwikkel is nie, en dat die toets daarvan dus effe voortydig is. Die insluiting van 'n skryfkomponent in 'n nagraadse akademiese geletterdheidstoets is, in die lig van Butler (2009:296) se navorsing, egter byna 'n onvermydelike gegewe. In sy opname onder nagraadse studieleiers aan die UP bevind Butler (2009:297) dat akademici oor die algemeen van mening is dat die aktiwiteit waarmee nagraadse studente die meeste probleme ondervind akademiese skryf is. Indien TAGNaS dus nie 'n akademiese skryfoefening sou bevat nie, sou potensiële gebruikers van die toets, veral akademici, dit waarskynlik as 'n groot tekortkoming in die ontwerp van die toets beskou. Bachman en Palmer (1996:24) waarsku dat aspekte omtrent die outentisiteit van 'n toets nie ligtelik deur ontwerpers opgeneem moet word nie, onder andere as gevolg van “its potential effect on test takers’ perceptions of the test”.

Butler (2009:297) kom ook tot die gevolgtrekking dat 'n negatiewe persepsie soos die voorgenoemde die gesigsgeldigheid van 'n toets kan aantast wat, in terme van die teoretiese raamwerk waarop hierdie studie geskoei is, op 'n verontagsaming van die psigiese analogiese ontwerpmoment neerkom, wat inhou dat 'n toets intuïtief aanneemlik en aanvaarbaar, selfs aantreklik, vir die gebruikers daarvan moet wees. Davies et al. (1999:59) definieer gesigsgeldigheid as “[t]he degree to which a test appears to measure the knowledge or abilities it claims to measure, as judged by an untrained observer (such as the candidate taking the test or the institution which plans to administer it)”. Besorgdheid oor die gesigsgeldigheid van 'n toets word dikwels as onbeduidend afgemaak, aangesien dit meer met die uiterlike voorkoms van 'n ontwerp as die onderliggende konstruk wat gemeet word, te make het. Die argument hou egter ook steek dat versuim om kwessies rakende gesigsgeldigheid te oorweeg uiteindelik die publieke reputasie of geloofwaardigheid van 'n toets (geloofsanalogie) in gevaar kan stel en dat die aanneemlikheid van 'n toets (die

tegnies gestempelde psigiese analogie) ’n praktiese oorweging is wat toetsontwerpers nie kan bekostig om te ignoreer nie (Davies et al., 1999:59).

Wat die nasienprotokol vir hierdie afdeling van TAGNaS betref, word daar in die besonder op die navorsing van Pot (2013) en Pot en Weideman (2015) gesteun. In haar diagnostiese studie van TALPS bied Pot (2013) ’n verfynde weergawe van die merkrubriek aan wat oorspronklik vir die nasien van die skryfstuk in TALPS ontwerp is. Hierdie herontwerpte protokol is die resultaat van die kombinasie van twee rubrieke wat deur Pot (2013) in die uitvoering van haar studie gebruik is. Een daarvan is die oorspronklike rubriek waarna sopas verwys is. Hierdie rubriek bestaan in hoofsaak uit ’n evaluering van die verskillende dele van die opstel in terme van inhoud, tegniese aspekte en taalgebruik.

Die tweede rubriek bied ’n evaluering van die skryfstuk uitsluitlik in terme van die komponente van die konstruk van akademiese geletterdheid wat in noue verband staan met analitiese onderskeiding as die bepalende kenmerk van akademiese diskoers, en wat van beduidende belang vir akademiese skryf is. Hierdie rubriek is spesifiek deur Pot (2013) vir haar studie ontwerp met die doel om, as ’n diagnostiese instrument, lig te probeer werp op die moeilikhede wat studente met betrekking tot akademiese geletterdheid en akademiese skryf ondervind. In kombinasie lewer hierdie twee rubrieke saam “[a] new marking rubric [...] which ensures – through an equal distribution of scoring points between content and language – a more extensive coverage of those aspects important to academic writing” (Pot, 2013:58). Vir hierdie verfynde nasienprotokol van Pot (2013) raadpleeg Bylaag C. Bylaag D bevat ’n tweede weergawe van hierdie protokol, vertaal en aangepas (waar nodig) vir nasiendoeleindes van die skryftaak in TAGNaS. Soos wat Pot (2013) se studie gewys het, sal ’n diagnostiese analise van die resultate van die skyfoefening, met sy rubriek, toetsontwikkelaars uiteindelik nader kan inlig omtrent watter probleme studente met akademiese skryf ondervind.

## 6. Die gebruik van veelkeuse-items in TAGNaS

Genoegsame regverdiging vir die gebruik van die veelkeusetoetstegniek vir akademiese geletterdheidstoetsing is alreeds by die ontwerp van TALL deur die ontwikkelaars daarvan verskaf (Van Dyk & Weideman, 2004b). Die grootte van die toetspopulasies en die dringendheid waarmee die toetsresultate van hierdie toets destyds deur die UP bekend gemaak moes word, het die gebruik van die veelkeuseformaat genoodsaak. Die onontkombaarheid van die aanwending van hierdie tegniek het die ontwikkelaars van TALL gedwing om doelbewus verby die intuïtief negatiewe reaksies oor hierdie toetsmetode te kyk en eerder die potensiaal daarvan op kreatiewe wyses te ontgin. Algemene kritiek wat dikwels teen hierdie tegniek geopper word, is dat dit slegs herkenningvermoë en nie produksievermoë nie toets, dat dit beperkinge plaas op die tipe kennis wat getoets kan word omdat afleiers nie altyd beskikbaar is nie en dat dit baie moeilik is om goeie veelkeuseitems te ontwerp (Hughes, 2003:76-78). Van Dyk en Weideman (2004b:16) lewer egter bewys van die teendeel van hierdie kritiek en meen dat

[i]n being forced to conceive of ways of employing this format, we have become more inventive and creative than we would otherwise have been, if we had simply succumbed to the prejudice that ‘one cannot test (this or that aspect of) language in this way’. The ingenuity of some of the questions has demonstrated to us that this bias is often exactly that: a prejudice that prevents us from seeking creative solutions.

Uit die enkele voorbeelditems wat reeds in hierdie hoofstuk aangebied is, behoort dit duidelik te wees hoedat daar ook in die ontwerp van TAGNaS gepoog is om op verbeeldingryke en kreatiewe wyse met die veelkeusetegniek om te gaan. Net soos met sy voorgangers, behoort die taaktipes in TAGNaS tot “a set of abilities [...] that are much broader in scope than that of a test that defines academic literacy in terms of skills, or reduces it to the mastery of sound, form and meaning” (Van Dyk & Weideman, 2004b:16). TAGNaS bevat nogtans, sáám met die veelkeuseafdelings, ook ’n akademiese skryfvraag wat pertinent op die assessering van kandidate se produktiewe skryfvermoë gerig is, ’n subtoets wat reeds hierbo in besonderhede uiteengesit en gemotiveer is.

Benewens die feit dat punttoekenning in 'n veelkeusetoets vinnig en maklik geskied, is 'n groot voordeel van dié formaat ook dat punttoekenning volkome betroubaar kan plaasvind. Verder maak hierdie formaat voorsiening vir die insluiting van meer toetsitems as wat dalk andersins moontlik sou wees omdat kandidate telkens slegs 'n merkie op papier hoef te maak sonder om antwoorde woord-vir-woord uit te skryf. Beide hierdie aspekte dra by tot groter toetsbetroubaarheid (Hughes, 2003:76). Wat die aantal opsies by elke vraag betref, word daar gewoonlik aanbeveel dat minstens vier opsies in elke geval verskaf moet word om die effek van korrekte raaiskote op toetsresultate te verminder (Hughes, 2003:77). In die geval van TAGNaS (wat boonop ook nog op nagraadse studente gemik is), word daar slegs van veelkeuseitems met vier of vyf opsies gebruik gemaak.

## **7. Die keuse van 'n tema en leestekste vir TAGNaS**

Die lees van tekste neem in verskeie definisies van akademiese geletterdheid 'n belangrike plek in. Soos vroeër aangetoon is, beskou Blanton (1994:8) interaksie met tekste as die kern van akademiese lees- en skryfgedrag. Spack (1997:3) definieer akademiese geletterdheid as die vermoë om voorgeskrewe tekste op universiteit te kan lees en skryf, terwyl Johns (1997:34) onder meer doeltreffende leesvermoë en goeie skryfvaardigheid as kernbeginsels van akademiese geletterdheid identifiseer. So ook Lea en Street (1998:160) wat lees en skryf as tipiese akademiese geletterdheidspraktyke bestempel waardeur studente kennis omtrent hulle onderskeie vakdissiplines opdoen. In hulle verfynde konstruk van akademiese geletterdheid lys Patterson en Weideman (2013b) “met tekste in wisselwerking tree” as een van die belangrike komponente van hierdie vermoë.

Die identifikasie en insluiting van geskikte tekste in 'n akademiese geletterdheidstoets is dus, in die lig van bogenoemde, 'n belangrike oorweging in die toetsontwerpproses. Net soos met TALPS het die tekste wat in TAGNaS gebruik word almal op dieselfde tema betrekking. By die seleksie van tekste is

daar, in die eerste plek, sorg gedra dat die tekste mekaar tematies komplementeer, aangesien die tekste tesame voldoende inligting moet verskaf vir gebruik in die uitvoering van die akademiese skryftaak in die derde en laaste gedeelte van die toets. Die tema van biodiversiteit en bewaring is vir die doeleindes van hierdie studie gekies. Aangesien hierdie onderwerp van aktuele belang is, behoort die meeste nagraadse studente in mindere of meerdere mate met hierdie tema vertrouwd te wees, sonder dat die inligting in die tekste egter as algemene kennis bestempel sou kon word. Volgens Du Plessis (2012:60) versterk tekste met aktuele onderwerpe waaroor gereeld in die openbare domein debat gevoer word die belyning van die toetskonstruk met outentieke diskoers, 'n oorweging wat strook met die vereistes van kommunikatiewe taaltoetsing en 'n verruimde siening omtrent taal. Hughes (2003:14) beveel ook aan dat tekste gekies moet word wat kandidate sal interesseer, maar nie sal ontstem of ontstignie, en dat tekste verkieslik nie te kultuurgebonde moet wees nie. Kontroversiële onderwerpe soos godsdiens, ras, misdad, siekte en persoonlike gesondheid moet in die reël vermy word. Onderwerpe van dié aard kan negatiewe emosies by kandidate ontlok wat 'n nadelige uitwerking op toetsprestasie kan hê en sodoende die regverdigheid van 'n toets kompromitteer (veral van belang wanneer ons die analogies juridiese momente in die tegniese metingsproses oorweeg).

Omdat TAGNaS nie as 'n vakspesifieke akademiese geletterdheidstoets bedoel is nie, word daar in hierdie studie van die standpunt uitgegaan dat vakinhoud en -terminologie nog nie in die stadium waarop die toets afgelê gaan word deur applikante vir nagraadse studie in 'n toets van akademiese geletterdheid getoets hoef te word nie. Daar is daarom toegesien dat die inhoud van die tekste wat vir TAGNaS geselekteer is, nie só dissipline-spesifiek is dat kandidate (uit watter vakgebiede ook al) dit buitengewoon moeilik sal vind om te begryp of daarby aansluiting te vind nie. Ooreenkomstig die tekste in TALPS het die beoogde tekste vir TAGNaS egter 'n besliste intellektueel-akademiese inslag. Die hoofteks vir die teksbegripsubtoets is byvoorbeeld uit 'n welbekende populêr-

akademiese tydskrif verkry en vertaal, terwyl die bykomende teks waarop die vergelyking-tussen-tekste-gedeelte gebaseer is, uit 'n resente wetenskaplik-toeganklike akademiese bron verkry is – beide tekste verteenwoordig skryfwerk wat 'n genoegsame intellektuele uitdaging aan nagraadse studente sal bied, sonder dat vakspesifieke kennis 'n voorvereiste vir begrip is. Du Plessis (2012:59-60) gaan trouens van die hipotese uit dat

[...] electing to use subject content of which test takers have a general, but not in-depth knowledge, will have the effect of compelling test takers to draw on language-related abilities, which may provide a better indication of literacy levels than would be the case when employing subject material from the test takers' own or chosen fields of study.

Aangesien TAGNaS vir voornemende nagraadse studente ontwerp word, moet die tekste in die toets van 'n toepaslike moeilikheidsgraad wees. Van Dyk (2010:218) verskaf die volgende matriks op grond waarvan die geskiktheid van 'n teks vir insluiting in 'n toets bepaal kan word:

<b>Kommentaar op teks</b>			
Leesbaarheidsindeks	Graadvlak	Lengte van die teks	Kommentaar ter verbetering (indien toepaslik)
bv. 55%	bv. Graad 11	bv. 500 woorde	bv. Te min diskoersmerkers

*Tabel 3.8 Matriks vir beoordelaars se kommentaar oor 'n teks*

Die eerste leesteks wat vir gebruik in die teksbegripsubtoets van TAGNaS geselekteer is, beskik oor 'n Flesch Kincaid-leesbaarheidsindeks van 34.7 en 'n graadvlak van 13.3. Hierdie is 'n vertaalde teks waarvan die oorspronklike Engelse weergawe 'n leesbaarheidsindeks van 51.3 en 'n graadvlak van 10.4 het. Omdat TAGNaS op nagraadse studente gemik word, moes die vertaalde weergawe van hierdie teks 'n meer uitdagende reproduksie van die oorspronklike wees, en geskikte aanpassings is dienooreenkomstig tydens die vertaalproses aangebring. Die Flesch Kincaid-leesbaarheidsformule bepaal die

moeilikhedgraad van 'n teks aan die hand van 'n leesbaarheidstelling tussen 0 en 100. Hierdie telling word op grond van bepaalde wiskundige berekeninge aan 'n teks toegeken, tesame met 'n graadvlak wat aan die toegekende telling gekoppel word. Gemiddelde sinslengte en die gemiddelde aantal lettergrepe per woord word onder andere by hierdie formule betrek. Hoe makliker 'n teks is, hoe hoër is die leesbaarheidstelling uit 100, met 50 as 'n gemiddelde waarde. 'n Telling van tussen 50 en 60 korreleer met graadvlakke 10 tot 12 op skool, terwyl 'n leesbaarheidsindeks van 0 tot 30 gewoonlik vir gegradueerde persone geskik is. Dus, in terme van leesbaarheid en graadvlak, kan die tellings wat hierbo vir die leesteks in TAGNaS aangegee word as gepas vir studente op nagraadse vlak beskou word. Hierdie tellings strook ook met dié van die leesteks in TALPS wat 'n leesbaarheidsindeks van 36.4 het (Rambiritch, 2012:156). Wat die lengte van hierdie teks betref, is dit 626 woorde lank, maar word aangevul met 'n tweede korter teks van 159 woorde wat in vergelyking met die eerste een gelees moet word. Dus 'n woordtotaal van 785 woorde, 'n lengte wat aanvaarbaar is volgens die nagraadse toetsstandaard van 600-800 woorde vir 'n leesteks. Hoewel die tweede leesteks uit minder woorde as die eerste een bestaan, is die moeilikhedgraad daarvan op 'n hoër peil met 'n Flesch Kincaid-leesbaarheidsindeks van 20.

Daar bestaan tans geen leesbaarheidsformule wat uitsluitlik vir Afrikaanse tekste uitgewerk is nie. Hierdie gegewe belemmer die maak van vergelykende uitsprake oor die noukeurigheid waarmee 'n internasionale formule soos Flesch Kincaid spesifiek Afrikaanse tekste se moeilikhedsvlak meet. Selfs al is die Flesch-formule "the best known readability formula in current use" (Fulcher, 1997:500) het leesbaarheidsformules in alle geval beperkte voorspellingswaarde. Faktore soos bv. swak grammatikale, kontekstuele en konseptuele struktuur wat nie deur tradisionele leesbaarheidsformules verreken word nie, kan 'n invloed op die leesbaarheid van 'n teks hê. Fulcher (1997) beveel daarom aan dat besluite omtrent die leesbaarheid van 'n teks nie slegs op die telling van 'n leesbaarheidsindeks gebaseer moet word nie, maar ook op deskundige oordeel.



Soos uit die voorlaaste afdeling van hierdie hoofstuk duidelik sal word, het 'n ervare ontwerpspan waarvan verskeie deskundiges deel was met die ontwikkeling van loodstoetsitems vir TAGNaS gehelp. Hulle insette is, ook wat die keuse van tekste betref, deurgaans tydens die ontwerp van TAGNaS in ag geneem en verreken.

## **8. Die woordeskatitems in TAGNaS en die *Akademiese Woordelys vir Afrikaans***

TAGNaS bevat, net soos sy Engelse en Afrikaanse voorgangers, 'n subtoets wat spesifiek kandidate se kennis van algemene akademiese woordeskat toets (kyk toetsspesifikasies, Bylaag B). Die woordeskatlys waarop hierdie subtoets in TAGNaS gebaseer word, is, soos reeds vantevore vermeld, Steyn (2013) se *Akademiese Woordelys vir Afrikaans* wat, as deel van die navorsing vir haar doktorsale proefskrif, as 'n Afrikaanse teenhanger vir Coxhead (2000) se AWL en die Nederlandse *Woordenlijst Academisch Nederlands* geskep is. Steyn (2013) dui aan dat die *Akademiese Woordelys vir Afrikaans* voortgebring is volgens dieselfde metodologieë wat in die samestelling van die voornoemde twee woordeskatlyste gebruik is. Steyn (2013) verduidelik dat Coxhead se lys saamgestel is uit 'n korpus van ongeveer 3.5 miljoen woorde wat uit 'n steekproef van tekste verkry is oor 28 vakgebiede heen binne die vier dissiplines van die kunste, handel, regte en natuurwetenskap. Die *Woordenlijst Academisch Nederlands* se steekproef het bestaan uit 1 410 871 akademiese tekste wat naslaanwerke, sillabusse en eerste- en tweedejaarstudiegidse van die Radboud Universiteit Nijmegen ingesluit het, asook 3 731 336 vaktydskrif- en koerantartikels. Net soos die *Akademiese Woordelys vir Afrikaans*, is beide hierdie lyse daarop gemik om woorde te identifiseer wat oor die algemeen in akademiese tekste verskyn, eerder as om die terminologie van verskillende vakgebiede aan te dui.

Vir die samestelling van die Akademiese Woordelys vir Afrikaans het Steyn (2013) 'n totale korpus van meer as 21 miljoen woorde gebruik, wat verkry is uit 'n wye versameling van Afrikaanse akademiese tekste soos vaktydskrif- en koerantartikels (spesifiek die *By* supplement van koerante soos *Volksblad*, *Beeld* en *Die Burger*), naslaanwerke en studiemateriaal. 'n Elektroniese frekwensielys is op grond van hierdie datastel gegenereer wat na afloop van 'n sistematiese seleksie- en verfyningsproses tot 'n Afrikaanse akademiese woordeskatlys van 2000 items gereduseer is. Die woorde in hierdie lys kan óf alfabeties óf volgens frekwensie gerangskik word, en word net soos in Coxhead se AWL in terme van woordfamilies gegroepeer, bv. die woorde *band(e)*, *bindend(e)*, *gebind*, *gebonde* en *verband* is almal morfologies verwant aan die werkwoord *bind*.

Die seleksie van woorde uit Steyn (2013) se woordeskatlys wat in TAGNaS getoets word, is dus nie bloot 'n willekeurige seleksie nie, maar is op grond van frekwensiekategorie bepaal. Omdat hierdie toets vir voornemende nagraadse studente bedoel is, is daar pertinent toegesien dat die grootste persentasie van antwoorde uit die lae frekwensiekategorieë (kategorie 8-10) van dié lys geneem word. 'n Bevredegende getal woorde uit die middel-frekwensiekategorieë (kategorie 5-7) word ook gebruik, terwyl die kleinste gedeelte van die seleksie uit hoëfrekwensiewoorde (kategorie 1-4) bestaan. Die insluiting van hoëfrekwensiewoorde in 'n nagraadse woordeskatsubtoets is steeds belangrik, aangesien items met woordopsies wat waarskynlik meer bekend aan kandidate sal wees, moontlik as 'makliker' inleidende items tot die subtoets ingespan kan word. Sodanige items help om te verseker dat kandidate nie met die intrapslag reeds deur die toetsitems in 'n bepaalde afdeling ontmoedig of ontsenu word nie. In die volgende hoofstuk word daar in die vorm van statistiese analise oor die loodsing en produktiwiteit van die woordeskatitems verslag gegee.

## **9. Die skryf van die items in TAGNaS**

Die skryf van die items vir TAGNaS is deur 'n span van ontwikkelaars behartig wat almal óf tans verbonde is aan óf voorheen geaffilieer was met drie van ICELDA se vennootuniversiteite, naamlik die UV, die UP en die NWU. Etlike van die lede van hierdie span het ruime ondervinding van taaltoetsontwikkeling op tersiêre vlak deurdat hulle reeds aan verskeie soortgelyke navorsingsprojekte van ICELDA deelgeneem het. Die navorser was dus verseker dat die toetsontwikkelaars wat by hierdie studie betrek is oor voldoende kennis en bewese praktiese ontwerpervaring van akademiese geletterdheidstoetse beskik. Hoewel sekere van die toetsitems reeds gedurende November en Desember van 2015 ontwikkel is, is die meerderheid daarvan tydens 'n eenmalige ontwerpbyeenkoms in Januarie 2016 geskep. 'n Program en inligtingstuk is vir hierdie byeenkoms saamgestel en spesifieke ontwerptake is aan die onderskeie spanlede toegewys. Sorgvuldige oriëntering, veral met betrekking tot die konstruk en toetsspesifikasies, is aan al die spanlede verskaf alvorens daar met die ontwerp van die items begin is. Die doelwit wat aanvanklik gestel is, was dat die loodstoetsweergawe van TAGNaS teen einde Januarie 2016 voltooi moes wees sodat loodstoetsing reeds vanaf die eerste semester van 2016 uitgevoer kon word. Hierdie oogmerk is suksesvol bereik en die loodsing van die konseptuële toets kon voortgaan soos beplan (kyk tabel 3.9 vir skedule).

Beoogde datum	Ontwerptaak	Voltooi teen	Kommentaar
31 Desember 2015	Skommeltekste Tekstipe Grammatika en teksverband Opsomming Verwysings	15 Desember 2015	Gedoen
15-17 Januarie 2016	Akademiese woordeskat Interpretasie van visuele inligting Teksbegrip Vergelyking tussen tekste	22 Januarie 2016	Gedoen
	Akademiese skryf	–	Uitgestel totdat loodstoetsing voltooi is
31 Januarie 2016	Finalisering van loodstoetsweergawe	31 Januarie 2016	Gedoen

*Tabel 3.9 Skedule vir die skryf van items vir TAGNaS*

## 10. Slot

Hierdie hoofstuk is geskryf met die oogmerk om ’n aanduiding te gee van hoe die eerste stappe in die toetsontwerpsiklus van ’n beoogde Afrikaanse nagraadse toets van akademiese geletterdheid verloop het (vgl. Fulcher, 2010:94). Daar is spesifiek gefokus op die artikulasie van ’n toetskonstruk en die belyning van hierdie konstruk met ’n bloudruk en ’n reeks meegaande toetsitems. Versuim deur toetsontwikkelaars om hierdie stappe uit te voer, kan soos Du Plessis (2012:66) dit stel “have serious repercussions in terms of construct validity and could even render any interpretations of ability based on the test scores to be null and void”. Die sorgvuldigheid waarmee ’n toets ontwerp en saamgestel word, maak uiteindelik die verskil uit wanneer toetse vir hulle kwaliteit vergelyk word, en lê die grondslag vir die soms moeisame verwerwing van ’n goeie reputasie.

Dit behoort duidelik uit hierdie hoofstuk te blyk hoedat sorgvuldige oorweging geskenk is aan die teoretiese rasionaal wat die grondslag vorm vir die seleksie en ontwerp van taaktipes in hierdie studie. Daar is gepoog om voldoende motivering te verskaf vir die sienswyses wat in hierdie studie gehuldig word

omtrent dit wat 'n nagraadse toets van akademiese geletterdheid behoort te toets, en op watter wyses sodanige assessering kan of behoort te geskied, met elemente van vernuwing daarby. In hierdie verband is bewuste pogings aangewend om die outentisiteit van die toets te verseker (vgl. Bachman en Palmer, 1996:23), met bewustelike erkenning aan analitiese onderskeiding as die bepalende kenmerk van akademiese diskoers. Nie net berus die toetskonstruk op 'n oop, kommunikatiewe siening van taal nie, maar korrespondeer die gekose subtoetse en taaktipes ook in 'n hoë mate met die tipe take wat studente buite die toetskonteks in die universitêre taalomgewing behoort uit te kan voer. Boonop maak outentieke tekste ook deel van die toetsontwerp uit.

Die hoofstuk wat volg, bied 'n oorsig van die resultate van die loodstoetsing (en die daaropvolgende verfyning) van die subtoetse en taaktipes. Daar word aan die hand van statistiese analises ondersoek ingestel na die produktiwiteit al dan nie van die onderskeie subtoetse en hulle individuele items.

## **Hoofstuk 4**

### **Die loodsing en verfyning van 'n nagraadse toets van akademiese geletterdheid in Afrikaans**

#### **1. Inleiding**

Bachman (2004:14-18) dui aan dat die meetproses in taaltoetsing oorhoofs uit drie logies geordende stappe bestaan. Die eerste stap behels die formulering van 'n konseptuele definisie van die toetskonstruk; die tweede stap vereis die daarstelling van 'n operasionele definisie van die konstruk in die vorm van toetsspesifikasies; en die derde stap bestaan uit die kwantifisering van die daaropvolgende waarnemings. Daar is in die voorafgaande hoofstuk pertinent aan die eersgenoemde twee sake aandag gegee, terwyl die fokus in die komende twee hoofstukke spesifiek op die derde stap val. Dit is volgens Bachman (2004:4) belangrik om te verstaan dat taaltoetsing 'n navorsingsveld is wat tegelyk uit twee wetenskaplike kontekste put. Eerstens is daar die toegepaste linguistiese konteks (vgl. Hoofstuk 2) wat aspekte dek soos die aard van taalgebruik, die aanleer van taal, taalvermoë en taalgebruikstake. Hierdie konteks lê die grondslag vir 'n definisie van die vermoë(ns) wat deur 'n taaltoets gemeet word, of, anders gestel, verskaf dit, in terme van die teoretiese onderbou van hierdie studie, 'n tegniese-analitiese funderende teoretiese rasionaal vir 'n taaltoets (vgl. Hoofstuk 3). Die tweede konteks wat ter sprake is, en wat in hierdie hoofstuk, sowel as die volgende een, die vernaamste wetenskaplike aandeel het, is die metingskonteks. Bachman (2004:4) wys daarop dat hierdie konteks hoofsaaklik met twee aspekte gemoeid is: “the relationship between the quantitative results of assessments on the one hand and their meaning, interpretation and use on the other”.

Omdat die resultate van taaltoetse gebruik word om afleidings oor kandidate se taalvermoë te maak en bepaalde besluite oor hulle te neem, moet toetsontwikkelaars kan demonstreer dat die resultate wat verkry word betroubaar

en die interpretasies daarvan geldig is. Om dus te verseker dat taaltoetse op 'n verantwoordelike manier gebruik word “we need to provide evidence that supports this use” (Bachman, 2004:4). 'n Belangrike tipe bewys wat versamel word om ondersteuning van dié aard te verskaf, word van kwantitatiewe data (d.i. toetsresultate), en die statistiese analises daarvan, afgelei. Die hoofdoel van hierdie hoofstuk is om sodanige bewyse vir die beoogde Toets van Akademiese Geletterdheid vir Nagraadse Studente (TAGNaS) aan te bied. In terme van die raamwerk van ontwerpbeginsels waarop hierdie studie berus, hou die voorgenome oogmerk spesifiek verband met die tegnies gestempelde numeriese ontwerpmoment wat voorhou dat, waar verantwoordelike ontwerp hoog op prys gestel word, veelvuldige bewyse op 'n sistematiese manier vir die doeleindes van toetsvalidering geïntegreer behoort te word.

Let op dat hierdie hoofstuk slegs die resultate van die eerste loodstoetsing van TAGNaS rapporteer, saam met besonderhede oor die daaropvolgende verfynings-/seleksieproses wat plaasgevind het. 'n Tweede rondte van loodstoetsing is egter ook onderneem waarvan die resultate eers in die volgende hoofstuk uiteengesit word.

## **2. Die toetsontwikkelingproses voortgesit: Die eerste loodstoetsing van TAGNaS**

### **2.1 Die nodigheid en belang van loodstoetsing**

Ná die ontwerp van die toetsspesifikasies en -items behels die volgende stap in die toetsontwerpproses die loodstoetsing en evaluering van die toetsitems (Fulcher, 2010:94). Bachman en Palmer (1996:91) verwys na hierdie stap as die proefstadium van die toetsadministrasiefase van toetsontwikkeling. Dit hou in dat 'n toets deur 'n verteenwoordigende steekproef van individue voltooi word ten einde inligting oor die bruikbaarheid daarvan te bekom, asook vir die doeleindes van moontlike verbetering en hersiening. Statistiese analises van die

data wat tydens loodstoetsing verkry word, verskaf inligting oor byvoorbeeld die eienskappe van elke toetsitem, soos moeilikheidsgraad en diskriminasiewaarde, asook oor die interne konsistensie (betroubaarheid) van 'n toets in sy geheel. Sodanige analyses fasiliteer daaropvolgende verfynings aan 'n toets wat byvoorbeeld die verwydering en/of herskryf van sekere items of hulle afleiers kan behels (Davies et al., 1999:213). Davies et al. (1999:213) dui aan dat die uiteindelijke oogmerk van loodstoetsing is om die kwaliteit van die toetsitems te verseker, asook die geskiktheid daarvan vir die beoogde doel van die toets en die toetspopulasie.

## 2.2 Getalle en deelnemers

Soos reeds in Hoofstuk 3 aangedui is, het die eerste loodsfase van die konseptoets van TAGNaS gedurende die eerste semester van 2016 plaasgevind. Die loodstoetspopulasie het bestaan uit ingeskrewe Afrikaanse studente aan al vier die vennootuniversiteite van die Inter-Institutional Centre for Language Development and Assessment (ICELDA), wat die Universiteit van die Vrystaat (UV), die Universiteit van Pretoria (UP), Universiteit Stellenbosch (US) en Noordwes-Universiteit (NWU) insluit. Die deelnemende studente was verteenwoordigend van 'n verskeidenheid fakulteite, departemente en studierigtings regoor die verskillende kampusse, onder andere akademiese geletterdheid, opvoedkunde, regte, arbeidsterapie, verpleegkunde, landbou, antropologie, Afrikaans en Nederlands, kommunikasiekunde, politieke wetenskap en drama en teaterkuns.

Aangesien 'n aansienlike groottotaal van 385 loodstoetsitems (kyk tabel 4.1) binne die tydsbestek van slegs 'n paar weke op toetspopulasies van bevredigende grootte uitgetoets moes word, is daar van 'n gemengde groep voorgraadse studente gebruik gemaak wat van eerste- tot derdejaarsvlak gewissel het. Om sake te bemoeilik, was klasse in die eerste kwartaal van 2016 vir 'n geruime tyd lank op verskeie kampusse landswyd ontwrig weens gewelddadige protesaksies



deur studente. Dit was dus prakties onmoontlik om 'n groot genoeg kohort van slegs nagraadse Afrikaanse studente vir die loodstoetsing byeen te kry wat boonop gewillig sou moes wees om meer as een loodstoetsessie by te woon vir die voltooiing van die verskillende subtoetse. Soos egter later duidelik sal word, is daar deurgaans met hierdie gegewe rekening gehou tydens die verfyning van die konseptuële toets. Omdat die uiteindelijke toets vir seniorvlak voorgraadse studente of voornemende nagraadse studente bedoel is, is daar sover moontlik by die seleksie van items voorkeur gegee aan produktiewe items met laer, maar steeds aanvaarbare fasiliteitswaardes (P-waardes), met ander woorde, items wat op grond van die statistiese analises as moeiliker of meer uitdagende items getoets het.

Die groot getal loodstoetsitems wat uitgetoets moes word, is hoofsaaklik toe te skryf daaraan dat 'n meerfasige toetsontwerp uiteindelik vir TAGNaS beplan word. Daar moes dus nie net vir die attrisie van onproduktiewe items voorsiening gemaak word nie, maar ook vir die samestelling van addisionele subtoetse wat (verdere) eksperimentering met 'n meerfasige toetsmodel moontlik sou maak. Boonop vergroot meer proefitems uiteindelik die kans op 'n meer betroubare meetinstrument.

Tabel 4.1 (kyk volgende bladsy) gee 'n uiteensetting van die aantal items wat oorspronklik per subtoets vir die loodstoetsing van TAGNaS ontwerp is, asook van die aantal items wat na afloop van die eerste data-analise- en verfyningsproses vir verdere gebruik geselekteer is. Van die 385 oorspronklike individuele toetsitems is 131 voorlopig vir die eerste twee verfynde toetsfases/dele van TAGNaS geselekteer. Daar word later in hierdie hoofstuk meer inligting verskaf oor die kriteria op grond waarvan die itemseleksie plaasgevind het (kyk afdeling 2.4), asook meer besonderhede oor die verfyningsproses self (kyk afdeling 4).

Die aantal kandidate wat per subtoets aan die eerste loodstoetsing deelgeneem het, word ook in tabel 4.1 verskaf. Die totale kohort het bestaan uit 3 601 kandidate, en, soos duidelik uit die tabel blyk, was geeneen van die toetspopulasies wat die verskillende subtoetse afgelê het kleiner as 129 kandidate per subtoets nie. Die grootste toetsgroep het uit 362 studente bestaan.

Subtoets/Taaktipe	Aantal loodstoetsitems (Ontwerp)	Grootte van loodstoetspopulasie	Aantal geselekteerde items (Verfyning)	
			Toets 1	Toets 2
<b>Skommeltekste</b>	<b>30</b>	<b>1 208</b>	<b>10</b>	<b>5</b>
Skommelteks 1	5	339	5	
Skommelteks 2	5	144	5	
Skommelteks 3	5	168		
Skommelteks 4	5	230		
Skommelteks 5	5	180		
Skommelteks 6	5	147		
<b>Interpretasie van visuele inligting</b>	<b>32</b>	<b>254</b>	nvt	<b>12</b>
<b>Akademie se woordeskaf</b>	<b>60</b>	<b>225</b>	<b>10</b>	<b>10</b>
Enkelwoord-items	40	225	8	8
Woordpaar-items	20	225	2	2
<b>Tekstipe</b>	<b>10</b>	<b>351</b>	nvt	<b>5</b>
Tekstipe 1	5	161		
Tekstipe 2	5	190		
<b>Teksbegrip (Vergelyking tussen tekste inkluis)</b>	<b>102</b>	<b>518</b>	nvt	<b>35</b>
Teksbegrip 1	34	175		
Teksbegrip 2	34	170		
Teksbegrip 3	34	173		
<b>Grammatika en teksverband</b>	<b>120</b>	<b>361</b>	<b>18</b>	<b>16</b>
Grammatika en teksverband 1	60	181		
Grammatika en teksverband 2	60	180		
<b>Opsomming</b>	<b>11</b>	<b>555</b>	nvt	<b>5</b>
Opsomming 1	5	362		
Opsomming 2	6	193		
<b>Verwysings</b>	<b>20</b>	<b>129</b>	nvt	<b>5</b>
Verwysings 1	10	129		
Verwysings 2	10	129		
<b>Totaal</b>			<b>38</b>	<b>93</b>
<b>Groottotaal</b>	<b>385 items</b>	<b>3 601 kandidate</b>	<b>131 items</b>	

Tabel 4.1 Eerste loodstoetsing van TAGNaS: 'n Oorsig van getalle (loodstoetsitems, -kandidate en geselekteerde items)

## 2.3 Procedure en benadering

Aangesien dit nie prakties haalbaar was om die totale getal ontwerpte toetsitems binne een enkele toetsessie op kandidate te loods nie, is hierdie items volgens subtoets in 'n reeks kleiner klastoetse verdeel wat elk binne die bestek van een periode deur kandidate afgelê kon word. Optiese leeskaarte wat elektronies per rekenaar by 'n dienspunt van die UV nagesien word, is as antwoordstelle vir hierdie toetse gebruik. Die uiteindelijke datastelle vir statistiese analise is saamgestel uit die toetsresultate wat in MS Excel-formaat van hierdie dienspunt ontvang is. Gedetailleerde data-analises is vervolgens met behulp van die IteMan 4.3- en TiaPlus-sagteware vir toets- en itemanalise uitgevoer. Beide hierdie programme genereer elektroniese data-analiseverslae gebaseer op klassieke toetsteorie (“classical test theory”). Die doel van hierdie verslae is om 'n aanduiding van die kwaliteit van die toetsitems te gee aan die hand van hulle psigometriese eienskappe (Guyer & Thompson, 2011:1). Vir die doeleindes van hierdie studie geld die verslae wat deur IteMan 4.3 gegenereer is as die primêre verslae vir die evaluering van toetsitems, terwyl die TiaPlus-verslae enkele addisionele statistiek en grafika verskaf wat telkens gebruik word om die IteMan-resultate te kruiskontroleer en te bevestig. Bykomend hiertoe, en slegs waar toepaslik, word enkele verwerkte data van die Winsteps-program ook by die bespreking van die toetsresultate betrek. Winsteps, anders as IteMan 4.3 en TiaPlus, assisteer met die toepassing van 'n spesifieke vertakking van item-responsteorie (“item response theory”) genaamd Rasch-analise.

Green (2013:xii-xiii) verduidelik dat daar oor die algemeen twee benaderings ten opsigte van die analise van toetsdata is, naamlik klassieke toetsteorie (KTT) en moderne toetsteorie, met laasgenoemde wat ook as item-responsteorie (IRT) bekend staan. KTT behels die analise van toetsdata ten einde aspekte te ondersoek soos die moeilikheidsgraad van items (fasiliteitswaarde), hoe goed items tussen kandidate diskrimineer (diskriminasiewaarde), hoe elke item tot die betroubaarheid van 'n toets bydra, die verhouding tussen die verskillende dele

van 'n toets en die verband tussen die eienskappe van kandidate en hulle toetsprestasie. Baie van hierdie analyses maak sterk staat op die korrelasiekoëffisiënt<sup>12</sup>.

KTT gaan van die veronderstelling uit dat omrede vermoëns (bv. taalvermoë) 'n abstrakte entiteit is dit nooit direk deur 'n toets gemeet kan word nie. In hierdie benadering word aanvaar dat 'n toetskandidaat se waargenome telling (“observed score”) die som van twee komponente is, naamlik die werklike telling (“true score”) plus 'n metingsfoutkomponent (“measurement error”), waar laasgenoemde dui op variansie wat nie vermoëverwant is nie, maar eerder onsistematies of willekeurig (“random error”) van aard. Hierdie aanname vorm die kern van KTT waaruit verskeie ander statistiek voortvloei wat deur programme soos Itean en Tiaplus bereken word (Cito, 2005). Davies et al. (1999:118) verduidelik dat “[e]rror is associated with all measurement of language, and interferes with the attempt to determine the true score of the test taker”. Hoewel ewekansige fout (“random error”) bereken kan word deur die gebruik van 'n statistiek bekend as die standaardmetingsfout (“standard error of measurement”), kan dit nie heeltemal uitgeskakel word nie en beïnvloed dit die betroubaarheid van 'n toets. Bronne van toetsontbetroubaarheid sluit bv. in variansie as gevolg van toetskandidate se gedrag (motivering, angs, ongesteldheid, raaiskote, ens.), faktore wat verband hou met die toetsituasie self (bv. probleme tydens toetsadministrasie), die invloed van agtergrondskennis op toetsprestasie, ens.

Die tweede benadering waarvan hierbo melding gemaak word, naamlik IRT, het ontstaan in 'n poging om in 'n bepaalde leemte van klassieke toetsanalise te voorsien. Alderson, Clapham en Wall (1995:89-92) verduidelik dat dit met KTT nie moontlik is om 'n vasgestelde meting van 'n toets se moeilikheidsgraad te verskaf nie, aangesien 'n toets se eienskappe in hierdie benadering nie van die eienskappe van die kandidate wat dit aflê, geskei kan word nie. Omdat die

---

12. 'n Korrelasiekoëffisiënt is 'n statistiek wat bereken word op grond van data wat die krag en rigting van die verhouding tussen twee veranderlikes opsom (Bachman, 2004:84).

fasiliteitswaarde van items die proporsie van kandidate verteenwoordig wat die items korrek beantwoord het, hang hierdie waarde onvermydelik af van die vermoë van die groep op wie die items uitgetoets word met die implikasie dat die resultate van die analyses in werklikheid slegs vir die betrokke steekproef geld en nie vir steekproewe van kandidate op verskillende vaardigheidsvlakke nie (Alderson et al., 1995:89). Dus is daar, soos McNamara (1996:153) dit stel “no way of knowing whether these characteristics of person ability and item difficulty would be maintained for the persons over different items and for the items if they were tried out on different subjects”. By KTT is daar daarom ’n sekere mate van onbestendigheid ter sprake: “traditional reports of candidate ability [...] and item characteristics [...] are potentially unstable [and] more susceptible to change if different test items and test subjects with differing levels of achievement are involved” (Davies et al., 1999:98).

IRT, en in die besonder Rasch-metingsteorie, oorbrug hierdie probleem deur die ontwikkeling van ’n moeilikheidsgraadskaal waarop items gerangskik word onafhanklik van die steekproef waarop dit geloods is (Alderson et al., 1995:90). Green (2013:xii) verduidelik dat IRT op waarskynlikheidsteorie gebaseer is; d.i. die waarskynlikheid van ’n gegewe persoon om ’n spesifieke item reg te beantwoord, word bepaal. Die basiese metode op grond waarvan hierdie waarskynlikheid bepaal word, kan soos volg toegelig word:

If we know how an individual has performed on other items (we have some estimate of his or her ability), and if we know how other individuals have performed on the item in question (we have an estimate of how difficult the item is), then we are in a position to make an estimate of how likely it is that the individual will respond in a particular way – that is, to choose either the ‘correct’ or the ‘incorrect’ answer. (McNamara, 1996:160)

Dus, indien ’n item maklik is en die kandidaat se vermoë goed is, is die waarskynlikheid hoog dat die item korrek beantwoord sal word. Insgelyks is die kans op ’n korrekte antwoord hoogs onwaarskynlik indien ’n item moeilik en die kandidaat se vermoë swak is. Met ander woorde, wat die Rasch-model doen, is om ’n wiskundige verband te lê tussen individuele vermoë en item-moeilikheidswaarde, en om hierdie verhouding dan in terme van

waarskynlikheid uit te druk (McNamara, 1999:153). Volgens Green (2013:xiii) maak IRT dit moontlik om, anders as KTT, sogenaamde steekproef-vrye moeilikheidswaarde (“sample-free item difficulty”) en item-vrye individuele vermoë (“item-free person ability”) te bepaal. McNamara (1999:153) beskou egter beskrywings soos hierdie as misleidend en onverstandig, aangesien alle estimasies van item-eienskappe, insluitende Rasch-estimasies, gebaseer word op response van bepaalde kandidate op bepaalde items, en sal uiteindelik beperk word deur die kwaliteit van die data waarop dit gebaseer word. Volgens hierdie outeur verskaf Rasch-analise “no magic solutions to problems of poor or inadequate data” (McNamara, 1996:154). Datastelle wat te klein is, lewer veral in hierdie benadering probleme op.

Trouens, die IRT-benadering is nie sonder struikelblokke van sy eie nie. Green (2013:xiii) lys verskeie redes waarom hierdie benadering nie voor die voet deur alle toetsontwikkelaars bo KTT verkies en gebruik word nie. Eerstens, vir IRT-analises om enigsins betekenisvol te wees, is ’n relatiewe groot aantal toetsresultate met ’n minimum van 200 kandidate, maar verkieslik meer as 300 individue per toets nodig – ’n vereiste wat nie altyd in die praktyk van taaltoetsing haalbaar is nie, veral nie by loodstoetsing nie. Daarbenewens was IRT-sagteware tot onlangs toe nog nie so geredelik beskikbaar soos KTT-programme nie, en word ook oor die algemeen as minder gebruikersvriendelik beskou. IRT-programme is byvoorbeeld gebaseer op die sogenaamde “log-odd unit scale” waarmee die meeste taaltoetsontwikkelaars nie vertrouwd is nie.

In hierdie hoofstuk word die toetsdata primêr vanuit die KTT-benadering geanaliseer en beoordeel. Slegs waar moontlik en van toepassing, word metings wat met behulp van Rasch-analise verkry is by die bespreking betrek. Die Rasch-analises wat wel ter sprake kom, word in die meeste gevalle bloot aangewend as verdere ondersteuning vir die interpretasies en gevolgtrekkings waartoe reeds gekom is op grond van die Iteman- en TiaPlus-berekeninge.

## 2.4 Produktiwiteit en parameters

Van Dyk en Weideman (2004b:17) verduidelik dat die uiteindelijke seleksie van items vir 'n spesifieke toets nie net afhang van die mate waarin die items met die toetskonstruk en -spesifikasies belyning is nie (kyk Hoofstuk 3), maar ook van die oordeel van die toetsontwerper wat betref hulle produktiwiteit. Hierdie oordeel kan gevel word op grond van óf toetsontwerpers se ondervinding en verwagtinge (kwalitatief), óf kwantitatiewe data oor die items, óf 'n kombinasie van beide. Van Dyk en Weideman (2004b:18) meen in hierdie verband dat toetsitems in terme van hulle beoordeling in vier kategorieë verdeel kan word, naamlik *aanvaarbaar* ('n hoë mate van belyning met die toetskonstruk, maar skynbaar nie produktief nie), *onaanvaarbaar* (lae produktiwiteit en 'n lae mate van belyning met die toetskonstruk), *wenslik* (produktief sowel as 'n hoë mate van belyning met die toetskonstruk), en *nie ideaal nie* (potensieel produktief, maar nie goeie belyning met die toetskonstruk nie). Tabel 4.2 bevat 'n visuele voorstelling van hierdie kategorisering (oorgeneem en vertaal uit Van Dyk & Weideman, 2004b:18).

<b>Belyning</b>	hoog	<i>aanvaarbaar</i>	<i>wenslik</i>
	laag	<i>onaanvaarbaar</i>	<i>nie ideaal nie</i>
		laag	hoog
		<b>Produktiwiteit</b>	

Tabel 4.2 'n Matriks vir die beoordeling van toetsitems

Indien toetsitems se produktiwiteit op grond van kwantitatiewe metings bepaal word, behels dit, in hierdie geval hoofsaaklik vanuit die oogpunt van KTT, dat psigometriese eienskappe van items soos die volgende – ooreenkomstig die parameters wat dit reguleer – deur toetsontwikkelaars in oorweging geneem word:

### *Die fasiliteitswaarde van items*

Soos reeds vantevore vermeld is, verwys die term fasiliteitswaarde na die moeilikheidsgraad van 'n item en word dit statisties as die P-waarde van 'n item uitgedruk. Die P-waarde gee proporsioneel 'n aanduiding van die aantal kandidate wat 'n item korrek beantwoord het. Dit kan van 0.0 tot 1.0 wissel. 'n Hoë waarde beteken dat 'n item maklik(er) is, terwyl 'n lae waarde daarop dui dat 'n item moeilik(er) is (Guyer & Thompson, 2011:29). Items wat uitermate maklik of moeilik is, word gewoonlik verwyder omdat dit, wat diskriminasiewaarde betref, geen bydrae tot 'n toets lewer nie. Trouens, items met 'n P-waarde van 0.95 of hoër diskrimineer gewoonlik baie swak tussen kandidate (Guyer & Thompson, 2011:29). Vir die doeleindes van hierdie studie is die afsnyppunte vir die minimum en maksimum P-waardes van geskikte toetsitems onderskeidelik 0.15 en 0.84. Hierdie afsnyppunte kom ooreen met die parameters wat ook by die ontwerp van verskeie ander akademiese geletterdheidstoetse van ICELDA as maatstaf gegeld het.

Volgens Davies et al. (1999:95-96) mik gestandaardiseerde toetse gewoonlik na 'n reeks items met 'n matige moeilikheidsgraadverspreiding binne die grense van 0.3 en 0.7, met 'n gemiddelde waarde van 0.5. Vir veelkeusetoetse word die moeilikheidsgraadindeks egter verhoog ten einde te kompenseer vir die waarskynlikheid dat kandidate sommige van die antwoorde reg kan raai. Omdat TAGNaS op seniorvlak studente gemik is, is daar, soos reeds aangedui, sover moontlik by die sifting van items voorkeur gegee aan produktiewe items met laer, maar steeds aanvaarbare P-waardes. In aansluiting hierby merk Green (2013:27) in hierdie verband op dat die geskiktheid van fasiliteitswaardes oorweeg moet word in die lig van die doel van die toets, die teikenpopulasie en hoe die waardes gebruik gaan word. Die insluiting van items met hoër P-waardes is egter ook noodsaaklik gegewe die algemene gebruik dat “[t]ests tend to begin with easy items in order to boost confidence and to ensure that weaker candidates do not waste valuable time on items which are too difficult for them” (Davies et al., 1999:96).



### *Die diskriminasiewaarde van items*

Die diskriminasiewaarde van 'n item gee 'n aanduiding van die mate waarin die item tussen sterker en swakke kandidate onderskei. In KTT is daar verskeie statistiese tegnieke wat gebruik kan word om itemdiskriminasie te bereken. In die geval van analyses uitgevoer deur Iteman 4.3 word die Pearson-punttweedelingskorrelasie (r-pbis-korrelasie) as 'n diskriminasiemeting gebruik waardeur itemresponse met toetspunttotele in verband gebring word. Die punttweedelingskorrelasie of r-pbis-waarde kan van  $-1.0$  tot  $1.0$  wissel. Wat die TiaPlus-program betref, word die Pearson-produk-moment-korrelasie (Rit-waarde) as 'n diskriminasiemeting gebruik. Die Rit-waarde kan van  $-1.0$  tot  $+1.0$  wissel (Cito, 2005:35).

'n Goeie item is daartoe in staat om tussen sterk en swak kandidate te onderskei en sal 'n hoër r-pbis-waarde hê, maar selde hoër as  $0.50$ . 'n Negatiewe r-pbis-waarde daarenteen is aanduidend van 'n baie swak item. Dit beteken dat die sterk kandidate binne 'n kohort 'n item verkeerd antwoord, terwyl die swak kandidate dit korrek beantwoord (Guyer & Thompson, 2011:30). Dit is wel gewens dat die afleiers van veelkeuseitems lae of negatiewe diskriminasiewaardes vertoon, aangesien slegs die korrekte antwoord veronderstel is om tussen kandidate te onderskei (Davies et al., 1999:97). Die minimum-diskriminasiewaarde wat as aanvaarbaar vir 'n item beskou kan word, is tipies 'n laerige, positiewe getal soos  $0.10$  of  $0.20$  (Guyer & Thompson, 2011:30). In die geval van hierdie studie is die minimumparameter vir itemdiskriminasie  $0.15$  met 'n maksimum wat minder as  $1.0$  sal wees, maar so hoog as moontlik. 'n Hoë r-pbis-waarde beteken dat kandidate met hoë toetspunttotele geneig is om die item korrek te beantwoord, terwyl kandidate met lae toetspunttotele geneig is om dit verkeerd te kry. Sodanige item word beskou as 'n item wat 'konsistent' met die toets is (Cito, 2005:35) en, volgens Davies et al. (1999:96), kan 'n toets met konsekwente hoë vlakke van itemdiskriminasie as betroubaar beskou word: "if these items have the capacity to spread individuals of differing ability along a single scale they must be repeatedly measuring the same underlying ability".

### *Betroubaarheid of interne konsistensie*

Betroubaarheid dui op die stabiliteit en konsistensie van 'n toets (Green, 2014:63-74; Van Dyk, 2010:128). Stabiliteit wys heen na die reproduseerbaarheid van toetsresultate wat, soos Fulcher en Davidson (2007:104) dit stel, daarop neerkom dat “[w]henever a test is administered, the test user would like some assurance that the results could be replicated if the same individuals were tested again under similar circumstances”. Konsistensie verwys weer na “die mate waartoe al die dele van 'n toets saamhang om dieselfde saak te meet” (Van Dyk, 2010:128). In terme van die teoretiese onderbou van hierdie studie is die vereiste dat toetsresultate toereikend konsekwent moet wees, ook met verloop van tyd, 'n tegnies gestempelde kinematiese analogie wat as 'n konstitutiewe ontwerpvoorwaarde geklassifiseer word.

Of 'n toets wel aan die bovermelde vereiste voldoen of nie kan op verskeie maniere bepaal word. Twee klassieke metodes is byvoorbeeld die *herhaalde toetsing* of *hertoetsing* van kandidate se prestasie tydens twee afsonderlike toetsgeleenthede (d.i. toets-hertoets-betroubaarheid), en/of die afneem van twee *parallele toetse* met soortgelyke maar verskillende vrae wat dieselfde konstruk meet (d.i. ekwivalente-vorm-betroubaarheid). Om via hierdie twee metodes by 'n betroubaarheidswaarde uit te kom, word 'n dubbele stel toetsgegevens dus benodig, waarvan die insameling egter 'n proses is wat verskeie unieke uitdagings bied en boonop baie tydrowend is. Dit het daarom algemene gebruik geword om 'n toets slegs een keer af te neem, en dan eerder te meet wat genoem word *inter-item-konsistensie* (Alderson et al., 1995:87). Davies et al. (1999:86) beskryf metings van interne konsistensie as effektiewe alternatiewe vir die voorgenoemde twee (meer omslagtige) benaderings.

Hierdie outeurs verduidelik hoedat statistiese metings of formules van interne konsistensie die nodigheid van hertoetsing uitskakel “by treating a single test as a number of smaller tests and comparing scores on the component parts” (Davies et al., 1999:86). Hulle beskryf toetsbetroubaarheid in hierdie opsig as “[t]he

actual level of agreement between the results of one test within itself”, en interne konsistensie as “[t]he degree to which scores on individual items or groups of items on a test correlate with one another” (Davies et al., 1999:86, 168). Die betroubaarheid van ’n toets word dus bereken deur te bepaal, soos Van Dyk (2010:154) dit verwoord, “of daardie items wat die konstruk (behoort te) verteenwoordig min of meer dieselfde resultate sal oplewer wanneer dit teen mekaar afgespeel word”. Volgens Van Dyk (2010:154) gaan interne konsistensie oor hoe stabiel kandidate se prestasie bly oor die verskillende dele/afdelings van ’n toets heen.

Binne KTT is daar ’n hele paar statistiese formules wat gebruik kan word om die interne konsistensie van ’n toets te bepaal. Op hierdie manier word die betroubaarheid van ’n toets in die vorm van ’n betroubaarheidskoeffisiënt gekwantifiseer. Een bekende metode is byvoorbeeld die halfverdelingsmetode (“split half reliability”) wat kortweg behels dat die items in ’n toets in twee dele verdeel word – ’n halfverdeling wat op verskillende maniere kan geskied – met die twee helftes wat as parallelle toetse hanteer en met mekaar gekorreleer word. Hoe sterker die twee helftes met mekaar korreleer, hoe meer betroubaar is die toets (Alderson et al., 1995:88). Nog ’n betroubaarheidsindeks, maar meer gesofistikeerd as die halfverdelingsmetode, is die Kuder-Richardson-formules (K-R20 en K-R21) wat die gemiddelde korrelasie tussen al die moontlike halfverdelings van die items in ’n toets gee. Hierdie formules word slegs gebruik by toetse wat digotomies van aard is (m.a.w. items kan slegs óf reg óf verkeerd wees). Volgens Davies et al. (1999:102) gee die K-R-formules ’n aanduiding van itemkonsistensie deurdat “they serve to determine whether candidates’ performance on any half of a given test is equivalent to performance on any other half”.

Die betroubaarheidskoeffisiënt wat deur beide die Iteman 4.3- en TiaPlus-program gebruik en gerapporteer word, is Cronbach se Alpha (aangedui met die simbool  $\alpha$ ). Die Alpha-waarde gee ’n aanduiding van hoe goed ’n groep items

saam die ter sake eienskap/vermoë meet deur proporsioneel te bepaal wat die toetsvariansie is as gevolg van gemeenskaplike faktore tussen die items onderling. Indien al die items in 'n toets dieselfde onderliggende vermoë meet, hoe sterker sal hierdie items met mekaar korreleer en hoe hoër sal die betroubaarheidswaarde van die toets wees. Net soos die Kuder-Richardson-formules is Cronbach se Alpha die gemiddelde van al die moontlike halfverdelingsbetroubaarheidskoeffisiënte, met die verskil dat die toetsitems nie noodwendig digotomies van aard hoef te wees nie (Davies et al., 1999:39). Cronbach se Alpha word gemeet op 'n skaal van -1 tot +1, met +1 wat aanduidend is van volmaakte interne betroubaarheid. Dus, hoe hoër die algehele Alpha-waarde van 'n toets is, hoe hoër is die interne konsistensie van die items. Volgens Green (2013:38) word 'n betroubaarheidskoeffisiënt van bo 0.7 as aanvaarbaar beskou, hoewel waardes van bo 0.8 verkies word. Hughes (2003:39) wys egter daarop dat betroubaarheid in die lig van die doel van 'n toets gesien moet word. Hoe belangriker die besluite is wat op grond van 'n toets geneem moet word, hoe groter moet die betroubaarheid wees wat vereis word. In die geval van TAGNaS wat as 'n Afrikaanse eweknie vir die Test of Academic Literacy for Postgraduate Students (TALPS) beoog word, is die uiteindelijke doelwit om 'n algehele betroubaarheidskoeffisiënt van bo 0.9 te bereik ten einde dieselfde hoë standaard te handhaaf wat deur die huidige weergawes van TALPS gestel word<sup>13</sup>. Dit sal weldra uit die toetsloodingdata duidelik word in watter mate TAGNaS (tot sover) in hierdie poging slaag al dan nie.

Buiten Cronbach se Alpha, bereken Iteman 4.3 ook drie vorme van halfverdelingsbetroubaarheid (naamlik “split-half random”, “split-half first-last” en “split-half odd-even”), eers as ongekorrigeerde korrelasies en daarna as gekorrigeerde Spearman-Brown-korrelasies (S-B). By die halfverdelingsmetode verteenwoordig die korrelasiekoeffisiënt wat bereken word die betroubaarheidswaarde van 'n toets wat helfte die lengte van die oorspronklike een is. Omdat toetsbetroubaarheid egter sterk deur die lengte van 'n toets

---

13. Verwys weer na tabel 3.6 in Hoofstuk 3 vir 'n uiteensetting van die betroubaarheidswaardes van die subtoets in TALPS soos afgelê in 2011 en 2012 by die UV.

beïnvloed kan word, moet die korrelasiekoëffisiënt gekorrigeer word ten einde by 'n betroubaarheidswaarde vir die totale lengte van die toets uit te kom. Die Spearman-Brown-formule word vir hierdie korreksie gebruik (Guyer & Thompson, 2011:23). Tog word vir die doeleindes van hierdie studie, waar Iteman 4.3-analises ter sprake kom, slegs Cronbach se Alpha as betroubaarheidsmeting verreken. Fulcher (2010:51) beskryf Cronbach se Alpha as “[t]he most frequently used and reported reliability coefficient”. TiaPlus bereken egter ook die “Greatest Lower Bound”-koëffisiënt (GLB) wat, volgens Van Dyk (2010:169), 'n meer akkurate skatting van betroubaarheid bied omdat Cronbach se Alpha, as gevolg van strengere parameters, inderwaarheid 'n onderskatting van die werklike betroubaarheid gee. Die ontwikkelaars van TiaPlus verduidelik dat die GLB “does not assume that splitting stops at the single item level. Grouping items iteratively, and by resampling, the GLB procedure is proven to find the greatest lower bound for test reliability” (Cito, 2005:37). GLB word veral gebruik vir toetse waarvan die konstruk multidimensioneel is (m.a.w. meer heterogeen), soos in die geval van hierdie studie, aangesien hierdie betroubaarheidsmeting nie homogeniteit veronderstel nie (Van der Slik & Weideman, 2005:26). Die vasstelling van die GLB-koëffisiënt maak gevolglik 'n belangrike onderdeel van die betroubaarheidsanalises in hierdie studie uit omdat, soos wat deur die *Standards for educational and psychological testing* vereis word, “[w]hen subtests of items within a test are dictated by the test specifications and can be presumed to measure partially independent traits or abilities, reliability estimation procedures should recognize the multifactor character of the instrument” (American Educational Research Association, 1999:33).

#### *Interne korrelasies en dimensionaliteit*

Davies et al. (1999:76) wys daarop dat interne betroubaarheidsmetings soos die halfverdelingsmetode, die Kuder-Richardson-formules en Cronbach se Alpha, in wese metings van toetshomogeniteit is. Homogeniteit is 'n eienskap van toetse wat te doen het met die ooreenkoms van die samestellende items of subtoetse

met mekaar in terme van dit wat gemeet word. Sodra die soeklig egter val op die eienskappe wat deur 'n toets gemeet word, kom die kwessie van konstrugeldigheid en -validering ook meteens aan die orde. Gevolglik het die mate waarin 'n toets homogeniteit weerspieël ook op die konstrugeldigheid daarvan betrekking, en as bewyse hiervoor dien onder andere gewoonlik inligting omtrent die korrelasiepatrone tussen die puntetellings van items en toetspunttotale (Bachman, 1990:258; Douglas, 2010:93). Ten einde bewyse rakende die homogeniteit van 'n toets te bekom, is 'n studie van die interne korrelasies daarvan dus nodig (Alderson et al., 1995:183-184), en vir sodanige ondersoek geld verskillende tipes korrelasiekoëffisiënte, elk met sy eie parameters (Van der Walt & Steyn, 2007:148). Wat die korrelasiekoëffisiënte tussen die verskillende subtoetse van 'n toets betref, word normaalweg aanbeveel dat dit van (om en by) 0.3 tot 0.5 moet wissel. Alderson et al. (1995:184) lig hierdie aanbeveling soos volg toe (vgl. ook Van der Walt & Steyn, 2007:147-148):

Since the reason for having different test components is that they all measure something different and therefore contribute to the overall picture of language ability attempted by the test, we should expect these correlations to be fairly low [...]. If two components correlate very highly with each other, say +.9, we might wonder whether the two subtests are indeed testing different traits or skills, or whether they are testing essentially the same thing.

As dit egter kom by die korrelasiekoëffisiënte tussen elke subtoets en die toets in sy geheel kan daar verwag word dat die korrelasies hoër sal wees, moontlik in die omgewing van +0.7 of meer, aangesien soos Alderson et al. (1995:184) dit stel, “the overall score is taken to be a more general measure of language ability than each individual component score”.

Verder kan faktoranalise ook gebruik word om te bepaal of 'n toets een- of multidimensioneel is, d.i. of een of meer vermoë(ns) tegelyk deur 'n toets gemeet word of nie. Indien 'n spreidingsdiagram aantoon dat 'n toets nie eendimensioneel is nie kan daar natuurlik geargumenteer word dat homogeniteit by die toets ontbreek. Oor hierdie kwessie moet daar egter wat hierdie studie betref enkele belangrike opmerkings gemaak word, veral in die lig daarvan dat

TAGNaS, net soos sy voorlopers (bv. TALL, TALPS), op 'n ryk geskakeerde konstruks van akademiese geletterdheid gebaseer is (kyk Hoofstuk 3).

Van der Slik en Weideman (2005:32) wys daarop dat in gevalle waar 'n multidimensionele toetskonstruk ter sprake is toetsontwikkelaars dit moontlik nodig mag vind om 'n tegniese kompromie te maak tussen die effektiwiteit van 'n absoluut homogene toets en die geskiktheid van 'n bepaalde taaktipe vir metingsdoeleindes – soos wat ook met TALL/TAG die geval was – waar by die ontwerp van dié toetse, argumente ten gunste van 'n ryk geskakeerde konstruks eerder as die bereiking van 'n volkome eendimensionele meetinstrument uiteindelik die swaarste geweeg het. Dit blyk dat 'n veelvlakkige konstruks soos akademiese geletterdheid onvermydelik met 'n sekere mate van heterogeniteit gepaard sal gaan. Hierdie outeurs voer derhalwe aan dat “[i]f some degree of heterogeneity is what is required to ensure contextual appropriateness, it may have to be tolerated” (Van der Slik & Weideman, 2005:32). Die waarskynlikheid bestaan dus dat ('n) soortgelyke kompromie(ë) in die geval van hierdie studie verdra sal moet word, 'n kwessie wat weer in die hieropvolgende paragrawe ter sprake sal kom.

Om saam te vat, berus die analise van die produktiwiteit van die loodstoetsitems in hierdie studie dus primêr op die volgende vier vrae en die parameters wat vir elkeen geld:

- Is die items geskik in terme van hulle fasiliteitswaarde? (P-waarde tussen 0.15 en 0.84)
- Is die diskriminasiewaarde van die items voldoende/bevredigend? (minimum r-pbis-waarde van 0.15, selde hoër as 0.50)
- Hoe goed is die betroubaarheidskoeffisiënt van elke subtoets/die toets in sy geheel? (Cronbach se Alpha en GLB)
- Is die interne korrelasies (bv. tussen die items/subtoetse onderling) bevredigend?

Omdat uitslae van Rasch-analise egter ook plek-plek in hierdie hoofstuk bykomend gebruik word tot staving van die produktiwiteit al dan nie van toetsitems, word kort agtergrond oor hierdie analisemodel vervolgens verskaf.

### *Enkele tersaaklike aspekte van Rasch-analise*

Green (2013:154-155) gee 'n bondige oorsig van hoe daar in Rasch-analise te werk gegaan word. Hierdie model neem beide die totale ruwe toetstellings van die toetskandidate en die moeilikheidsvlakke van die items (gebaseer op die aantal kandidate wat die items korrek beantwoord het) in aanmerking. Gebaseer op hierdie twee faktore, poog die model dan om die mees waarskynlike respons op elke item deur elke kandidaat te voorspel. Dit probeer ook om, gegewe die vermoëns van die toetskandidate, die waarskynlike moeilikheidsvlak van elke item te bepaal. Hierdie verwagte of voorspelde resultate word dan met die werklike of waargenome resultate vergelyk. Waar enige verskil tussen hierdie resultate groter is as wat geantisipeer is, word die analise outomaties weer uitgevoer. Die model gaan met ander woorde voort om die verhouding tussen die moeilikheidsvlak van die items en die vermoëns van die toetskandidate te analiseer totdat daar 'n bevredigende vlak van akkuraatheid ten opsigte van die data bereik is. McNamara (1996:162) verduidelik dat die estimasieprosedure as kalibrasie ('calibration') bekend staan en dat elke opeenvolgende, rekursiewe estimasiesiklus 'n iterasie ('iteration') genoem word. Wanneer die vereiste akkuraatheidsvlak vir 'n betrokke datastel bereik word, word dit as gekonvergeer ('converged') beskryf.

Indien die model nie daarin slaag om 'n bepaalde item of toetskandidaat in die verwagte raamwerk in te pas nie word sulke items of kandidate van verdere oorweging uitgesluit. Die model verwag byvoorbeeld nie dat 'n item 'n fasiliteitswaarde van 0 of 100% gaan hê, of dat 'n toetskandidaat al die items korrek of verkeerd gaan beantwoord nie. Indien dít gebeur, word sulke items en/of kandidate uit die analise uitgeskakel omdat die model in sodanige geval nie 'n akkurate prentjie van moeilikheidsvlak en/of vermoë kan verskaf nie.



Verder verwag die model ook dat sterk kandidate (soos bepaal deur hulle algehele prestasie op die toets) maklike items korrek gaan beantwoord, en dat swak kandidate moeilike items verkeerd gaan beantwoord. Waar dit nie gebeur nie, kan die items of kandidate geïdentifiseer word as onpassend met betrekking tot die verwagte responspatrone vanweë 'n gebrek aan konvergensie. McNamara (1996:169) beskryf hierdie aspek van Rasch-analise as “the question of fit” en onderskei drie gevalle van ‘fit’ wat in 'n datastel na vore kan kom:

- (1) those in which the pattern for individual items, allowing for normal variability, fits the overall pattern (items showing appropriate ‘fit’);
- (2) those in which the pattern for individual items does not correspond with the overall pattern (‘misfitting’ items); and
- (3) those in which the general pattern of responses, which shows some ‘noise’ or unpredictability [...] is not observed, and the lack of variability means that an essentially deterministic pattern emerges for the items concerned (‘overfitting’ items).

Die mate waarin items aanvaarbare vlakke van ‘fit’ oftewel passing vertoon al dan nie, word deur die Winsteps-program uitgedruk en opgesom in die vorm van “item fit statistics”. Daar sal in die loop van die hieropvolgende bespreking af en toe na hierdie stel statistiek vir sekere items verwys word.

Dit is ook belangrik om kennis te neem dat die maateenheid waarin die geraamde metings van moeilikheidsvlak (items) en vermoë (kandidate) in Rasch-analise uitgedruk word as “log odd units” of ‘logits’ bekend staan. Op die “log odd unit”-intervalskaal word die gemiddelde moeilikheidsvlak van items by nul logits vasgestel, terwyl bogemiddeld moeilike items 'n positiewe logit-waarde sal hê en makliker items 'n negatiewe een. Estimasies van kandidate se vermoë hou dan ook met hierdie estimasies verband. 'n Kandidaat met 'n vermoë-waarde van nul logits sal 'n 50% kans hê om 'n item van gemiddelde moeilikheidswaarde korrek te beantwoord. Sterker kandidate sal positiewe logit-waardes hê, terwyl negatiewe logit-waardes 'n aanduiding van swakker vermoë by kandidate is (McNamara, 1996:165-166).

Met die voorgaande inligting as agtergrond word die resultate van die statistiese analises van die eerste loodstoetsing van TAGNaS vervolgens aangebied en bespreek.

### **3. Die eerste loodstoetsing van TAGNaS: Analise en bespreking van toetsresultate per subtoets**

Vir die eerste loodstoetsronde van TAGNaS is die totale getal ontwerpte items volgens die verskillende tipes subtoetse in 'n hele aantal kleiner klastoetse verdeel. Hierdie verdeling is om verskeie praktiese redes gedoen waaroor reeds in 'n vorige afdeling van hierdie hoofstuk uitgebrei is. 'n Gevolg hiervan is dat die kwantitatiewe ontleding van die toetsresultate aanvanklik (d.i. vóór verfyning/seleksie) ook volgens subtoets uitgevoer moes word wat verklaar waarom die statistiese analises in hierdie afdeling op sodanige wyse ingedeel en bespreek word. Die rapportering van subtoetsdata word trouens ten sterkte aanbeveel, aangesien

[i]t is not sufficient to report estimates of reliabilities and standard errors of measurement only for total scores when sub scores are also interpreted. The form-to-form and day-to-day consistency of total scores on a test may be acceptably high, yet sub scores may have unacceptably low reliability. For all scores to be interpreted, users should be supplied with reliability data in enough detail to judge whether scores are precise enough for the users' intended interpretations. (American Educational Research Association, 1999:31)

#### **3.1 Skommeltekste**

Ses skommeltekste (bestaande uit vyf items elk) waarvan drie uiteindelik na afloop van die verfynings-/seleksieproses in die finale weergawe van TAGNaS ingesluit moet word, is aanvanklik vir die eerste loodstoetsronde ontwerp.

Tabel 4.3 verskaf 'n opsomming van die beskrywende statistiek vir die resultate van die individuele skommeltekssassesserings. Soos duidelik uit tabel 4.3 blyk, beskik al ses die skommeltekste oor hoogs bevredigende diskriminasiewaardes ("Gem. Rpbis"). Trouens vyf van die ses skommeltekste se

punttweedelingskorrelasies is hoër as 0.50, resultate wat, in die lig van die parameters wat vroeër in hierdie hoofstuk uiteengesit is, meer as tevrede stel (vgl. Guyer & Thompson, 2011:30). Wat die moeilikheidsgraad van die skommeltekste betref, val die P-waardes (“Gem. P”) van al ses die assesserings binne die geldende parameters van 0.15 en 0.84. Skommelteks 5 met ’n fasiliteitswaarde van 0.61 is die maklikste skommelteks in die reeks. Hoewel dié waarde nogtans nie uitermate hoog is nie en hierdie skommelteks steeds goed tussen kandidate blyk te diskrimineer (r-pbis van 0.52), sal dit hoogs waarskynlik nie by die verfynde weergawe van TAGNaS ingesluit word nie, aangesien daar in hierdie studie voorkeur gegee word aan items met laer, maar steeds aanvaarbare P-waardes, gegewe die teikenpopulasie van die onderhawige toets.

Subtoets	Items	Gem.	SA	Min. Telling	Maks. telling	Gem. P	Gem. Rpbis
Skommelteks 1	5	2.20	1.82	0	5	0.44	0.60
Skommelteks 2	5	2.63	1.92	0	5	0.53	0.64
Skommelteks 3	5	2.48	1.77	0	5	0.50	0.57
Skommelteks 4	5	1.71	1.41	0	5	0.34	0.42
Skommelteks 5	5	3.06	1.53	0	5	0.61	0.52
Skommelteks 6	5	2.75	1.78	0	5	0.55	0.55

*Tabel 4.3 Beskrywende statistiek vir skommeltekste soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)*

Tabel 4.4 hieronder bevat ’n betroubaarheidsanalise van dieselfde ses skommeltekste wat hierbo bespreek word. Op grond van die Alpha-waardes in die tweede kolom van hierdie tabel kan afgelei word dat Skommelteks 1, 2 en 3 meer betroubare assesserings as Skommelteks 4, 5 en 6 is. Daar is reeds vroeër aangetoon dat ’n betroubaarheidskoëffisiënt van 0.80 of hoër verkieslik is, alhoewel ’n waarde van bo 0.7 ook as toereikend geag word. Volgens dié maatstaf is slegs die Alpha-waarde van Skommelteks 4 nie van die vereiste gehalte nie. Hoewel hierdie inligting nie in tabel 4.4 vertoon word nie, bereken TiaPlus die GLB-koëffisiënt van die eerste drie skommeltekste onderskeidelik as 0.87, 0.90 en 0.89 – betroubaarheidswaardes wat by uitstek voldoen aan die maatstawwe waarvolgens daar in hierdie studie gemeet word.

Subtoets	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Rando m	S-B First- Last	S-B Odd- Even
Skommelteks 1	0.81	0.79	0.78	0.57	0.76	0.88	0.73	0.87
Skommelteks 2	0.84	0.77	0.68	0.62	0.79	0.81	0.77	0.88
Skommelteks 3	0.80	0.80	0.74	0.38	0.80	0.85	0.55	0.89
Skommelteks 4	0.66	0.83	0.31	0.52	0.44	0.47	0.69	0.61
Skommelteks 5	0.75	0.77	0.54	0.26	0.71	0.70	0.41	0.83
Skommelteks 6	0.78	0.83	0.66	0.66	0.66	0.80	0.79	0.79

Tabel 4.4 *Betroubaarheidskoëffisiënte van skommeltekste soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)*

Verder weerspieël subtoetse soos Skommelteks 1 en 2, wat die twee hoogste betroubaarheidsmetings van die groep het, voldoende homogeniteit in terme van die vermoë(ns) wat gemeet word (kyk bylaag B vir toetsspesifikasies). Die interne korrelasies tussen die items in hierdie subtoetse is relatief hoog met die oorgrote deel van die waardes telkens bo 0.6 of 0.7 (kyk tabel 4.5 en 4.6).

Item	1	2	3	4	5
1					
2	0.7520				
3	0.4314	0.7560			
4	0.7097	0.7110	0.6885		
5	0.5996	0.6157	0.6927	0.8807	

Tabel 4.5 *Interne korrelasies tussen die items van Skommelteks 1 soos bepaal deur TiaPlus (Eerste loodstoetsing)*

Item	1	2	3	4	5
1					
2	0.7857				
3	0.7247	0.8906			
4	0.5822	0.7722	0.7445		
5	0.4483	0.5999	0.7131	0.9040	

Tabel 4.6 *Interne korrelasies tussen die items van Skommelteks 2 soos bepaal deur TiaPlus (Eerste loodstoetsing)*

Omdat Skommelteks 1 op 'n groter kohort van studente geloods is as die ander skommeltekste kon ook betekenisvolle “item fit”-statistiek met Rasch-analise vir hierdie subtoets verkry word. Die mate van passing (‘fit’) vir elke item tussen die voorspelde en werklike resultate word in tabel 4.7 verskaf met behulp van 'n

statistiek genaamd “Infit Mean Square (MNSQ)”. Hierdie statistiek gee ’n aanduiding van die mate van vertroue wat gestel kan word in die logit-meting (‘Measure’) wat aan elke item gekoppel word. Volgens Green (2013:169) is die ideale MNSQ-waarde vir ’n item gelyk aan 1.0, maar ’n waarde van tussen 0.5 en 1.5 word steeds as produktief beskou. ’n MNSQ-waarde van hoër as 1.5 word as onproduktief geag, aangesien sulke items te onvoorspelbaar ten opsigte van die model optree en dus onpassend of ‘misfitting’ is. Net so word ’n MNSQ-waarde van onder 0.5 ook as onproduktief bestempel omdat sodanige items op hulle beurt weer te voorspelbaar optree (‘overfitting’), d.i. al die maklike items word reg beantwoord en die moeilike items verkeerd. In tabel 4.7 val al die items se “Infit MNSQ”-waardes binne die voorkeurparameters van 0.5 en 1.5 wat daarop dui dat hierdie items statisties optree soos verwag. Die logit-metings van die items wissel van -2.17 (Item 1) tot 1.20 (Item 4) en die metingsfoutwaarde van 0.17 tot 0.20 (“Model SE”). Hierdie foutwaardes is aanvaarbaar, aangesien die foutgrens vir produktiewe items lê tussen -0.3 en 0.3 (Green, 2013:169).

Nog ’n belangrike statistiek in tabel 4.7 is die “Pt Measure” wat aandui in watter mate toetskandidate se response met die moeilikheidsvlak van die items korreleer. Die CORR-statistiek beskryf die waargenome korrelasie en die EXP-statistiek die verwagte korrelasie. Die waardes in beide kolomme blyk bevredigend te wees, aangesien geen daarvan negatief of te naby aan nul (rondom 0.1) is nie. Wat diskriminasiewaarde betref, eggo die “Exact Match”-statistiek hieronder die positiewe Iteman 4.3-resultate wat vroeër gerapporteer is (kyk weer tabel 4.3). Die klein verskil tussen elke item se OBS- (waargenome) en EXP-persentasie (verwagte) by hierdie statistiek is aanduidend van goeie itemdiskriminasie (Green, 2013:170).

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S. E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM
3	138	339	.27	.17	1.11	1.4	1.25	2.0	A .73	.76	72.9	73.2	3
5	118	339	.92	.19	1.06	.7	.97	-.1	B .77	.77	79.3	78.9	5
1	224	339	-2.17	.19	1.05	.6	1.02	.2	C .69	.70	77.7	77.6	1
4	111	339	1.20	.20	.96	-.3	.78	-1.1	b .79	.78	83.0	82.6	4
2	155	339	-.21	.17	.84	-2.3	.81	-2.4	a .78	.75	75.5	70.6	2
MEAN	149.2	339.0	.00	.18	1.01	.0	.97	-.3			77.7	76.6	
P. SD	40.5	.0	1.19	.01	.09	1.3	.17	1.5			3.4	4.2	

Tabel 4.7 “Item fit”-statistiek vir die items van Skommelteks 1 soos bepaal deur Winsteps (Eerste loodstoetsing)

### 3.2 Interpretasie van visuele inligting

Vir die loodstoetsing van hierdie subtoets is 32 items ontwerp met as mikpunt ’n seleksie van 12 geskikte items vir uiteindelijke gebruik. Iteman 4.3 stel die gemiddelde P-waarde van hierdie individuele subtoets vas op 0.51 met ’n gemiddelde diskriminasiewaarde van 0.30. Die algehele Alpha van die subtoets staan op ’n hoogs bevredigende waarde van 0.81; insgelyks die GLB-koëffisiënt wat as 0.91 aangedui word. Ondanks hierdie goeie resultate is daar egter op itemvlak enkele items wat tydens die analiseproses as probleemitems uitgewys is wat uiteindelik óf aan verfyning onderwerp óf bloot uit die verfynde weergawe van die toets weggelaat moet word. Tabel 4.8 bevat ’n lys van vyf items wat deur Iteman 4.3 as probleemitems (“flagged items”) geïdentifiseer of ‘uitgewys’ is (sien in dié verband ook tabel 4.25), waar tabel 4.9 die itemstatistiek van elk van hierdie items weergee.

Item ID	P / Item Gem.	R	Uitwysing
1	0.886	0.136	HP, LR
4	0.154	0.097	LR
23	0.110	-0.044	K, LP, LR
24	0.248	-0.026	K, LR
29	0.516	0.112	LR

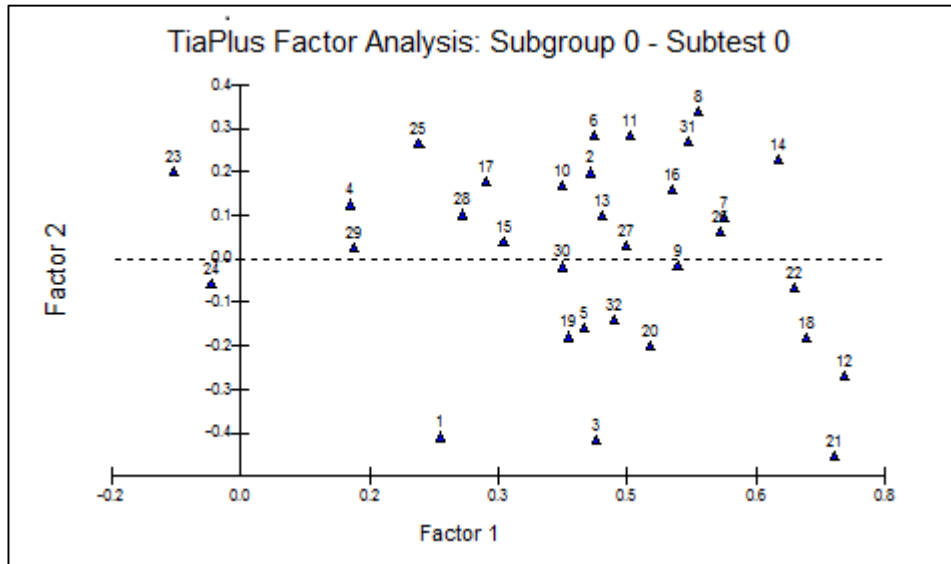
Tabel 4.8 Geïdentifiseerde probleemitems: Interpretasie van visuele inligting (Eerste loodstoetsing)

Item ID	N	P	Totale Rpbis	Totale Rbis	Alpha w/o
1	254	0.89	0.14	0.22	0.81
4	254	0.15	0.10	0.15	0.81
23	254	0.11	-0.04	-0.07	0.82
24	254	0.25	-0.03	-0.04	0.82
29	254	0.52	0.11	0.14	0.81

*Tabel 4.9 Itemstatistiek van probleemitems: Interpretasie van visuele inligting (Eerste loodstoetsing)*

Uit die bostaande twee tabelle is dit duidelik dat Item 1 met 'n P-waarde van 0.89 'n uitermate maklike item is wat gevolglik redelik swak tussen kandidate diskrimineer (rpbis = 0.14). Item 4 en 29 se diskriminasiewaardes is ook te laag en val buite die parameters wat vir hierdie studie gestel is. Item 23 en 24 blyk hoogs problematiese items te wees met negatiewe diskriminasiewaardes wat indien dit uit die subtoets weggelaat sou word, weliswaar 'n hoër betroubaarheidswaarde vir die geheel tot gevolg sou hê ("Alpha w/o" = 0.82).

'n TiaPlus-faktoranalise van hierdie subtoets bevestig tot 'n groot mate die bostaande bevindinge (kyk volgende bladsy). Soos Green (2013:122) dit stel, is 'n faktoranalise nuttig "for highlighting items which are performing differently – a confirmation if you like of what we find when carrying out a reliability analysis". Uit figuur 4.1 blyk dit duidelik hoedat veral Items 23 en 24, en in mindere mate Items 1, 4 en 29, uitskieters is wat nie saam met die res van die items in min of meer dieselfde domein kluster nie. Hoewel nie as probleemitems geïdentifiseer nie, kan die homogeniteit van Items 3 en 21 wat desgelyks weg van die groter groepering lê, moontlik ook bevraagteken word. Hier moet egter nie uit die oog verloor word nie dat die geskakeerdheid van die toetskonstruk 'n mate van heterogeniteit kan meebring waaraan speelruimte in die toetsontwerp gelaat behoort te word.



*Figuur 4.1 Faktoranalise: Interpretasie van visuele inligting (Eerste loodstoetsing)*

Rasch-analise verskaf addisionele bewyse ter bevestiging van die bevindinge omtrent die produktiwiteit van hierdie subtoets. Die konvergensietabel by 4.10 wys dat die aanvanklike kalibrasie van hierdie subtoets se datapatroon (“Prox iteration”) slegs twee iterasies vereis het en dat geen toetskandidate of -items van verdere analise uitgesluit is nie. Al 254 kandidate is behou, en so ook al 32 items. Die tweede deel van die estimasieprosedure (“JMLE iteration”) wat gebruik word om meer presiese estimasies te verskaf, het deur slegs vyf iterasies gegaan om die vereiste akkuraatheidsvlak vir die betrokke datastel te bereik. Item 24, wat ook deur Iteman 4.3 as ’n probleemitem uitgewys is, is geïdentifiseer as die item in die datastel wat die minste konvergerend ten opsigte van die verwagte patroon is.

PROX ITERATION	ACTIVE COUNT PERSON	ACTIVE COUNT ITEM	ACTIVE COUNT CATS	EXTREME 5 PERSON	RANGE ITEM	MAX LOGIT MEASURES	CHANGE STRUCTURE
1	254	32	2	3.73	2.76	2.2687	
2	254	32	2	4.21	3.08	.2945	

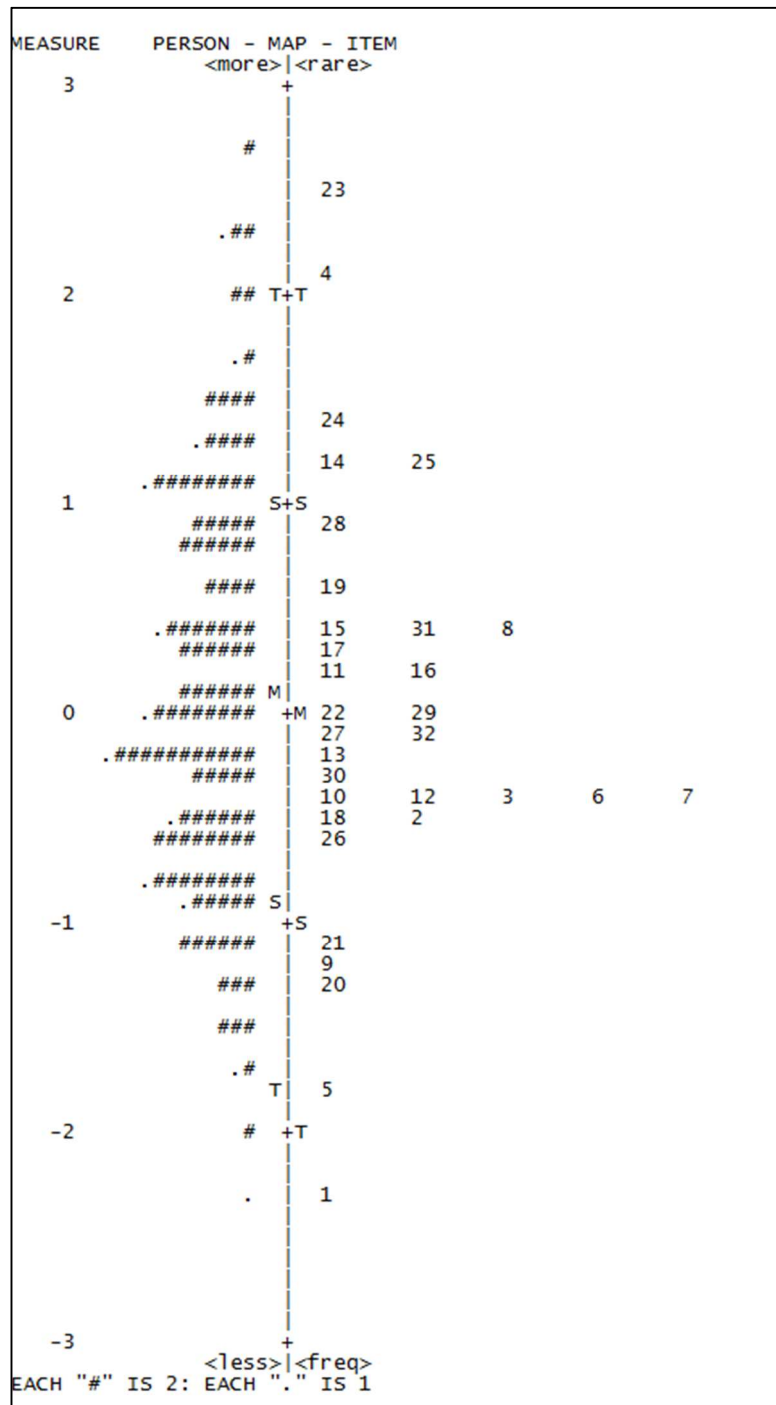
*Tabel 4.10 Konvergensietabel (Deel 1): Interpretasie van visuele inligting (Eerste loodstoetsing)*



JMLE ITERATION	MAX SCORE RESIDUAL*	MAX LOGIT CHANGE	LEAST PERSON	CONVERGED ITEM	CAT	CATEGORY RESIDUAL	STRUCTURE CHANGE
1	-3.11	.0554	28	24*			
2	-1.46	.0298	28	24*			
3	-.67	.0150	28	24*			
4	-.32	.0076	16	24*			
5	-.15	.0039	16	24*			

Tabel 4.10 *Konvergensietabel (Deel 2): Interpretasie van visuele inligting (Eerste loodstoetsing)*

Trouens, byna al die items wat deur Iteman 4.3 as problematies uitgewys is, word ook deur Rasch-analise so bevestig, soos blyk uit die onderstaande persoon-item-passing (“variable map”) wat deur Winsteps gegenereer is. Die item-estimasies wissel van ongeveer -2.3 logits (Item 1) tot 2.5 logits (Item 23). Items 4 en 23, asook Item 1, lê egter ver verspreid van die res van die items, d.i. anderkant die boonste en onderste tweede standaardafwyking (T), wat daarop dui dat hierdie items nie by die res van die items inpas nie.



Figuur 4.2 Persoon-item-passing: Interpretasie van visuele inligting (Eerste loodstoetsing)

### 3.3 Akademiese woordeskat

Vir hierdie subtoets is twee veelkeuse-itemtipes ontwerp waarvan die insluiting reeds in Hoofstuk 3 gemotiveer en bespreek is. In kort vereis die eerste, oorspronklike itemtipe dat slegs 'n enkele gaping in die leidraadsin met 'n

korrekte woordopsie voltooi moet word, terwyl die tweede, gewysigde itemtipe vereis dat 'n *dubbele gaping* tegelykertyd met twee korrekte opsies gevul moet word. Die uiteindelijke doel hiermee is die geregverdigde insluiting van beide itemtipes in die eerste én tweede toetsfase van TAGNaS. Tabel 4.11 verskaf 'n opsomming van die beskrywende statistiek vir die resultate van beide itemtipes in die woordeskatsubtoets.

Subtoets	Items	Gem.	SA	Min. telling	Maks. telling	Gem. P	Gem. Rpbis
Akademiese woordeskat (Enkele gaping)	40	26.22	4.77	12	37	0.66	0.20
Akademiese woordeskat (Dubbele gaping)	20	13.24	3.28	3	19	0.66	0.29

*Tabel 4.11* Beskrywende statistiek vir die akademiese woordeskatsubtoets soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)

Beide itemtipes se gemiddelde fasiliteits- en diskriminasiewaardes val binne die statistiese perke wat vir hierdie studie geld. Dit is opmerklik dat die gemiddelde diskriminasiewaarde van die gewysigde “dubbele gaping”-items effens hoër is as dié van die enkelwoorditems, ondanks die feit dat die gemiddelde P-waarde van beide subtoetse dieselfde is. Alhoewel daar nog nie in hierdie vroeë stadium van loodstoetsing tot enige finale gevolgtrekkings oor itemproduktiwiteit gekom kan word nie, wil dit tog voorlopig lyk asof eksperimentering met hierdie gewysigde taaktipe wel vrugte mag afwerp. Hierdie bevinding, hoewel steeds preliminêr, word verder bevestig deur die betroubaarheidswaarde wat vir hierdie taaktipe behaal is, soos verskaf in tabel 4.12.

Subtoets	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Akademiese woordeskate (Enkele gaping)	0.71	2.58	0.58	0.48	0.55	0.74	0.65	0.71
Akademiese woordeskate (Dubbele gaping)	0.71	1.76	0.53	0.54	0.55	0.69	0.71	0.71

*Tabel 4.12 Betroubaarkheidskoeffisiënte van die akademiese woordeskate-taaktipes soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)*

Trouens, albei die woordeskate-taaktipes hierbo toon 'n aanvaarbare Cronbach se Alpha van 0.71. Deur middel van die Spearman-Brown-korreksieformule stel TiaPlus egter die Alpha-waarde van die “dubbele gaping”-taaktipe nóg hoër vas op 'n waarde van 0.83, wat wel die geval sou wees indien hierdie subtoets langer was en uit ten minste 40 items, in plaas van slegs die huidige 20, bestaan het. Boonop is die GLB-koeffisiënt van beide taaktipes ook hoogs bevredigend en word dit onderskeidelik op 0.88 en 0.83 bereken.

Rasch-analise stel die itembetroubaarkheid van die enkel- en dubbel-gaping-subtoetse onderskeidelik vas op 0.97 en 0.98. Daar bestaan dus redelike sekerheid dat hierdie items op dieselfde manier sal optree indien dit op 'n ander toetspopulasie as die onderhawige een uitgetoets sou word. Die huidige populasie wat gebruik is vir hierdie subtoets het uit 225 toetslinge bestaan.

### 3.4 Tekstipe

Twee weergawes van hierdie subtoets (bestaande uit vyf items elk) is tydens die eerste loodstoetsing van TAGNaS op die proef gestel. Die grootte van die populasie was 161 studente. Die gemiddelde fasiliteitswaarde van Weergawe 1 word aangedui as 0.65 en die totale diskriminasiewaarde as 0.39. Vir Weergawe 2 staan hierdie selfde waardes onderskeidelik op 0.76 en 0.48. Wat interne betroubaarkheid betref, word die Alpha-koeffisiënt van die eerste weergawe bereken as 0.64 en dié van die tweede weergawe as 0.72.

Op grond van die bovermelde syfers is dit duidelik dat, hoewel beide proefweergawes se diskriminasiewaardes bevredigend is, Weergawe 1 se Alpha-betroubaarheidskoeffisiënt effens onder die geldende maatstaf van 0.7 lê<sup>14</sup>, terwyl Weergawe 2 se gemiddelde fasiliteitswaarde weer, met inagneming van die teikenpopulasie van die toets, te naby aan die maksimum perk van 0.84 is. In die geheel beskou, skiet beide weergawes dus in 'n bepaalde opsig tekort. Sodra die fokus egter van die geheel na die afsonderlike items binne elke weergawe verskuif word, raak dit gou duidelik dat beide weergawes wel oor individuele items met hoogs bevredigende statistiese eienskappe beskik. Die verfyning van hierdie subtoets sal daarom heel waarskynlik neerkom op 'n kombinasie van die individuele produktiewe items soos geïdentifiseer in beide van die proefweergawes. Tabel 4.13 en 4.14 bevat die beskrywende statistiek van die onderskeie individuele items in beide weergawes.

Item ID	N	P	Totale Rpbis	Totale Rbis	Alpha w/o
1	161	0.38	0.56	0.71	0.49
2	161	0.65	0.37	0.48	0.59
3	161	0.50	0.43	0.54	0.56
4	161	0.78	0.30	0.42	0.62
5	161	0.92	0.30	0.55	0.62

*Tabel 4.13 Itemstatistiek van Tekstipe: Weergawe 1 (Eerste loodstoetsing)*

Item ID	N	P	Totale Rpbis	Totale Rbis	Alpha w/o
1	190	0.69	0.50	0.65	0.66
2	190	0.85	0.44	0.67	0.69
3	190	0.87	0.42	0.67	0.70
4	190	0.62	0.58	0.74	0.63
5	190	0.75	0.47	0.64	0.68

*Tabel 4.14 Itemstatistiek van Tekstipe: Weergawe 2 (Eerste loodstoetsing)*

---

14. Dit is wel nodig om daarop te wys dat indien die betrokke subtoets uit 40 items sou bestaan, die Spearman-Brown-korreksieformule die Alpha-koeffisiënt van Weergawe 1 op 'n aansienlik hoër waarde van 0.93 beraam. Boonop word die GLB-koeffisiënt van Weergawe 1 reeds op 0.78 vasgestel. Dus, hoewel die huidige Alpha-waarde effens onder die maatstaf van 0.7 lê, behoort hierdie subtoets nie sonder meer as onbetroubaar afgemaak te word nie. Verder is die maatstaf van 0.7 ook strenger geldend vir die algehele betroubaarheid van 'n toets in sy geheel as vir die betroubaarheid van elke afsonderlike subtoets waaruit die geheel saamgestel word.

Wat tabel 4.13 betref, blyk dit dat Items 1 tot 3 die produktiefste items van die betrokke groep is deurdat dit 'n aansienlike daling van die algehele Alpha-waarde tot gevolg sou hê indien enige daarvan uit die toets weggelaat sou word ("Alpha w/o"). Tog is veral Item 5 'n probleemitem met 'n té hoë P-waarde van 0.92. Volgens tabel 4.14 kan Items 1, 4 en 5 as produktiewe items geklassifiseer word, terwyl Items 2 en 3 ontoereikende items met uitermate hoë P-waardes is wat nie aan die gestelde vereistes van hierdie studie voldoen nie. Tesame is daar nietemin, in die lig van die twee stelle data hierbo, genoeg individuele produktiewe items wat in kombinasie met mekaar gebruik kan word om 'n Tekstipe-subtoets van voldoende gehalte saam te stel.

### 3.5 Teksbegrip en Vergelyking tussen tekste

Wat die Teksbegrip-subtoets betref, is 'n aansienlike totaal van 102 veelkeuseitems aanvanklik getoets, almal gebaseer op dieselfde teksmateriaal, waarvan ongeveer 15 items spesifiek op die nuut-bykomende "Vergelyking tussen tekste"-gedeelte (waaroor reeds in Hoofstuk 3 meer vertel is) betrekking het. Omdat die totale getal teksbegrip-items te veel vir insluiting in 'n enkele toetsvraestel was, is hierdie subtoets verpak en geloods in die vorm van drie afsonderlike klastoetse bestaande uit 34 items elk. Gevolglik is die statistiese ontledings dienooreenkomstig hierdie verdeling uitgevoer wat verklaar waarom die resultate van hierdie subtoets telkens in drie dele (Teksbegrip 1, 2 en 3) hieronder verskaf word. Die verfyning van hierdie subtoets sal neerkom op 'n seleksie van ongeveer 40 geskikte items (hierby ingesluit items wat vergelyking behels) uit al drie die loodstoetse wat afgelê is.

Tabel 4.15 gee 'n aanduiding van die gemiddelde fasiliteits- en diskriminasiewaardes van die Teksbegrip-subtoets in sy drie dele, terwyl tabel 4.16 die resultate van die drieledige betroubaarheidsanalise uiteensit.

Subtoets	Items	Gem.	SA	Min. telling	Maks. Telling	Gem. P	Gem. Rpbis
Teksbegrip 1	34	18.84	4.42	9	29	0.55	0.19
Teksbegrip 2	34	20.47	4.86	7	33	0.60	0.23
Teksbegrip 3	34	19.11	5.18	7	32	0.56	0.26

Tabel 4.15 *Beskrywende statistiek vir die Teksbegrip-subtoets soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)*

Subtoets	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Teksbegrip 1	0.66	2.57	0.52	0.39	0.47	0.68	0.56	0.64
Teksbegrip 2	0.72	2.56	0.55	0.55	0.62	0.71	0.71	0.77
Teksbegrip 3	0.76	2.53	0.59	0.51	0.67	0.74	0.68	0.80

Tabel 4.16 *Betroubaarheidskoëffisiënte van die Teksbegrip-loodstoetse soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)*

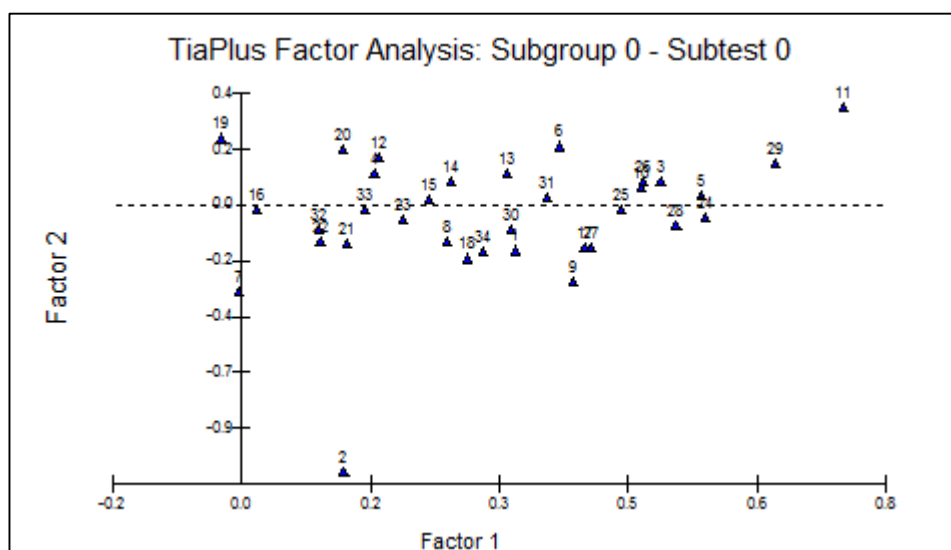
Van die drie loodstoetse wat geskryf is, lewer die items van Teksbegrip 3 die mees bevredigende resultate op met 'n algehele Alpha-koëffisiënt van 0.76 en 'n gemiddelde r-pbis-waarde van 0.26. Teksbegrip 1 se algehele Alpha-waarde is wel ietwat laer as 0.7, maar beskik nietemin oor 'n hoogs bevredigende GLB-koëffisiënt van 0.87 (kyk tabel 4.17). Trouens, die GLB-koëffisiënt van al drie die Teksbegrip-assesserings is hoër as 0.8, en die Rit-waarde telkens hoër as 0.25.

	Teksbegrip 1	Teksbegrip 2	Teksbegrip 3
<b>Number of persons in test</b>	175	170	173
<b>Average test score</b>	18.84	20.47	19.11
<b>Average P-value</b>	55.41	60.21	56.21
<b>Average Rit</b>	0.29	0.32	0.34
<b>Coefficient Alpha</b>	0.66	0.72	0.76
<b>GLB</b>	0.87	0.89	0.91
<b>Items used in GLB proc</b>	34	34	34
<b>Cut-off score</b>	10.5	10.5	10.5

Tabel 4.17 *Beskrywende statistiek vir die Teksbegrip-subtoets soos bepaal deur TiaPlus (Eerste loodstoetsing)*

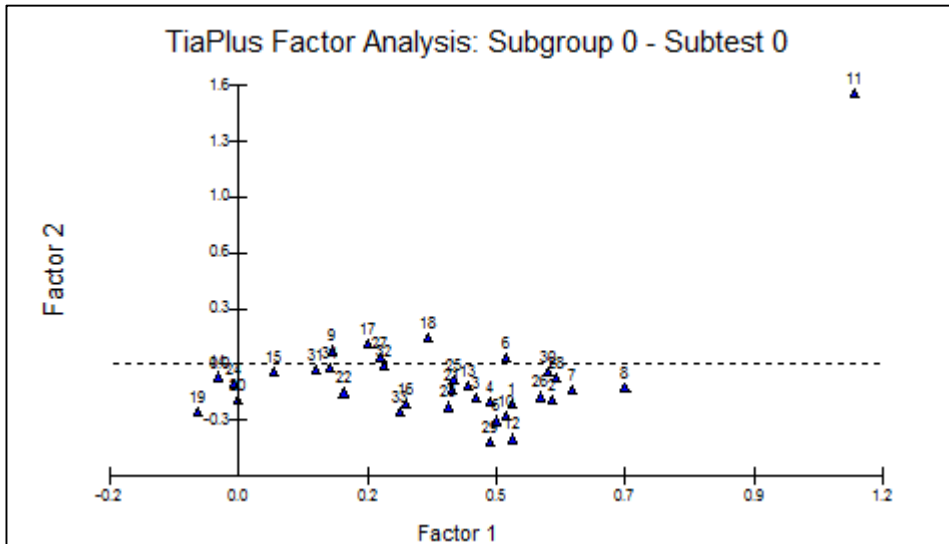
Wat dimensionaliteit betref, toon faktoranalises van die Teksbegrip-resultate hoedat, met uitsondering van enkele uitskieters, die meerderheid items van die

Teksbegrip-assesserings telkens in mindere of meerdere mate gekonsentreerd saam groepeer. Dit dien as bewys dat dié assesserings grotendeels eendimensioneel is, m.a.w. redelik homogeen in terme van dit wat gemeet word. Uitskieters by Teksbegrip 1 is Items 2, 7, 16 en 19 wat volgens Iteman 4.3 onproduktiewe items met onbevredigende diskriminasiewaardes is (kyk figuur 4.3). Teksbegrip 2 se spreidingsdiagram vertoon eweneens 'n digte groepering rondom die nullyn behalwe vir enkele uitskieters soos Items 11, 19 en 24 (kyk figuur 4.4). So ook vorm die meerderheid items van Teksbegrip 3 deel van die groter kluster in die regterhelfte van die diagram, hoewel 'n hele paar items ook verder weg van die groter groep na links uitstip (kyk figuur 4.5). Hierdie items is egter nie almal items wat statisties swak vaar nie. So byvoorbeeld wys die itemanalises van Items 2, 3, 23 en 26 dat hierdie items steeds aan die statistiese maatstawwe wat vir hierdie studie geld, voldoen. Die waarskynlikste rede waarom hierdie items, in terme van hulle homo-/heterogeniteit, effens anders as die res optree, is dat hulle 'n ander aspek of aspekte van teksbegrip meet – 'n subtoets waarvan die artikulasie met die komponente van die toetskonstruk in ieder geval opvallend geskakeerd is (kyk bylaag B vir toetsspesifikasies). Die mate van heterogeniteit wat deur items soos hierdie veroorsaak word, sou egter wel geduld kon word, aangesien dit nie uitermate groot blyk te wees nie.

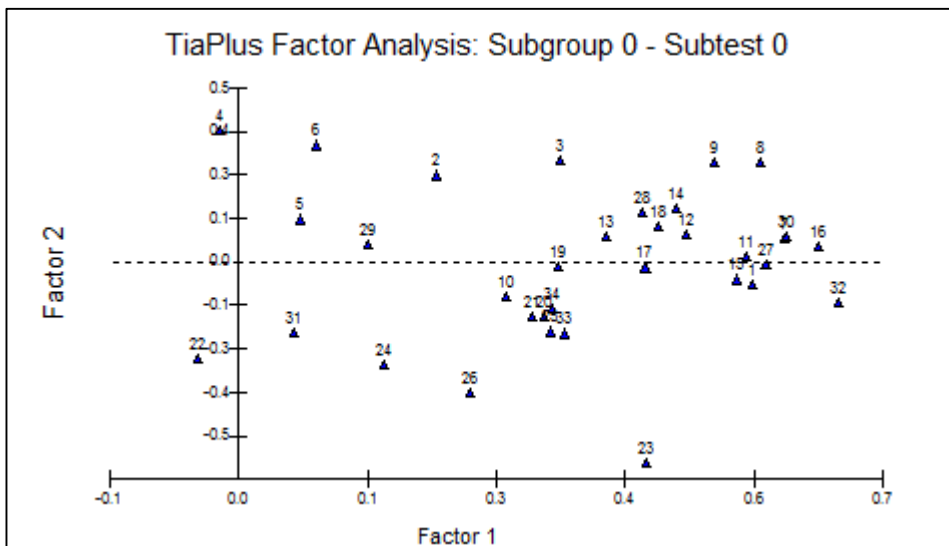


Figuur 4.3 Faktoranalise: Teksbegrip 1 (Eerste loodstoetsing)





*Figuur 4.4 Faktoranalise: Teksbegrip 2 (Eerste loodstoetsing)*



*Figuur 4.5 Faktoranalise: Teksbegrip 3 (Eerste loodstoetsing)*

Die itemanalises van die items wat spesifiek vergelyking en/of kontras tussen tekste behels, lewer grootliks bevredigende resultate op, met slegs enkele swak presterende items waarvan die diskriminasiewaardes (“Totale Rpbis”) buite die geldende parameters van hierdie studie val. Hierdie items word telkens in tabel 4.18 met ’n asterisk geïdentifiseer (kyk volgende bladsy). Die meerderheid oorblywende items in die tabel se statistiek is egter aanduidend van produktiewe items wat wel vir insluiting in die verfynde weergawe van TAGNaS oorweeg kan word. Dit wil dus ook in hierdie geval voorlopig lyk asof die vergelykende

teksbegrip-taaktipe 'n waardevolle nuwe toevoeging tot 'n toets van akademiese geletterdheid kan wees.

### 3.6 Grammatika en teksverband

Die oogmerk met hierdie subtoets is die insluiting daarvan in beide die eerste twee toetsfases van TAGNaS, 'n voorneme waarvoor die motivering reeds in Hoofstuk 3 verskaf is. Vier grammatika en teksverbandassesserings van 30 items elk is gevolglik in die eerste rondte van loodstoetsing op die proef gestel (dus 'n totaal van 120 items) met die doel om die twee produktiefste assesserings van die groep vir verdere verfyning en individuele itemseleksie te identifiseer. Die grootte van die populasies in hierdie geval was 181 en 180 studente. Volgens die inligting in tabel 4.19 beskik Grammatika en teksverband 1a en b tegelyk oor hoër gemiddelde r-pbis-waardes en laer gemiddelde P-waardes as 2a en b. Gegewe die geldende seleksiekriteria beteken dit dat 1a en b dus die twee mees geskikte weergawes van hierdie taaktipe vir verdere verfyning en gebruik is.

	Item ID	N	P	Totale Rpbis	Totale Rbis	Alpha w/o
Teksbeg. 1	30	175	0.33	0.19	0.25	0.65
	31	175	0.58	0.23	0.28	0.65
	32*	175	0.22	0.06	0.08	0.66
	33*	175	0.45	0.09	0.11	0.66
	34	175	0.41	0.20	0.26	0.65
Teksbeg. 2	31*	170	0.32	0.08	0.11	0.72
	32	170	0.74	0.16	0.21	0.72
	33	170	0.38	0.25	0.32	0.71
	34*	170	0.31	0.09	0.11	0.72
Teksbeg. 3	31*	173	0.27	0.05	0.07	0.77
	32	173	0.39	0.50	0.63	0.74
	33	173	0.41	0.24	0.31	0.76

Tabel 4.18 *Itemstatistiek van die vergelykende items in die Teksbegrip-loodstoetse (Eerste loodstoetsing)*

Subtoets	Items	Gem.	SA	Min. telling	Maks. telling	Gem. P	Gem. Rpbis
Grammatika en teksverband 1a	30	20.43	5.60	5	30	0.68	0.37
Grammatika en teksverband 1b	30	21.14	6.56	4	30	0.70	0.45
Grammatika en teksverband 2a	30	22.64	4.57	5	30	0.75	0.30
Grammatika en teksverband 2b	30	23.21	5.99	0	30	0.77	0.43

*Tabel 4.19 Beskrywende statistiek vir die Grammatika en teksverband-subtoets soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)*

Die geskiktheid van hierdie twee assesserings word verder bevestig deur hulle hoogs bevredigende individuele Alpha-waardes, onderskeidelik 0.85 en 0.90, en 'n gesamentlike Alpha-waarde van 0.93 (kyk tabel 4.20). Die GLB-koëffisiënt van elk word nóg hoër vasgestel op onderskeidelik 0.96 en 0.97 en die algehele GLB-waarde op 0.99. Die Alpha-waardes van Grammatika en teksverband 2a en b is effens laer as dié van 1a en b, maar nietemin steeds van 'n hoë standaard (kyk tabel 4.21).

Subtoets	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Items in totaal	0.93	3.03	0.89	0.69	0.89	0.94	0.82	0.94
Grammatika en teksverband 1a	0.85	2.15	0.73	0.61	0.81	0.84	0.75	0.89
Grammatika en teksverband 1b	0.90	2.07	0.80	0.64	0.86	0.89	0.78	0.92

*Tabel 4.20 Betroubaarheidskoëffisiënte van Grammatika en teksverband 1a en b soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)*

Subtoets	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Items in totaal	0.90	2.87	0.83	0.53	0.89	0.91	0.69	0.94
Grammatika en teksverband 2a	0.80	2.03	0.71	0.48	0.74	0.83	0.65	0.85
Grammatika en teksverband 2b	0.89	1.95	0.77	0.73	0.91	0.87	0.84	0.96

*Tabel 4.21 Betroubaarheidskoëffisiënte van Grammatika en teksverband 2a en b soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)*

### 3.7 Opsommingtaak

Hierdie subtoets wat die fokus op die visualisering van logiese onderskeidings laat val, is een van die vernuwende elemente waarmee daar in hierdie studie geëksperimenteer word. Twee verskillende weergawes daarvan is vir loodstoetsdoeleindes ontwerp. Die TiaPlus-analises in tabel 4.22 en 4.23 wys dat beide weergawes oor gemiddelde Rit-waardes van bo 0.5 beskik, asook oor P-waardes wat aanduidend is van 'n aanvaarbare moeilikheidsgraad. Ofskoon dit op die oog af lyk of die Alpha- en GLB-koëffisiënte van beide weergawes nie aan die statistiese verwagtings voldoen nie, moet daar in gedagte gehou word dat die kort lengte van hierdie subtoets 'n besliste negatiewe impak op die sterkte van die betroubaarheidswaardes het. Dié inwerking raak opmerklik sodra die Spearman-Brown-korreksieformule by die betroubaarheidsanalise in berekening gebring word, aangesien die geskatte Alpha-koëffisiënt van beide weergawes in sodanige geval aansienlik hoër is. Indien die subtoets uit 40 items sou bestaan, word die Alpha-waarde van Weergawe 1 as 0.89 aangedui en dié van Weergawe 2 as 0.85. Hierdie syfers in ag genome, kan daar in die onderhawige geval weereens met redelike sekerheid tot die voorlopige slotsom gekom word dat 'n subtoets van dié betrokke aard 'n waardevolle aanwinst vir 'n toets van akademiese geletterdheid kan wees.

<b>Number of persons in test</b>	362	<b>Number of selected items</b>	5
<b>Minimum test score</b>	0	<b>Maximum test score</b>	5
<b>Average test score</b>	2.39	<b>Standard deviation</b>	1.35
<b>Average P-value</b>	47.73	<b>Std. Error of Measurement</b>	0.94
<b>Average Rit</b>	0.59		
<b>Coefficient Alpha</b>	0.51	<b>SE Coeff. Alpha</b>	0.04
<b>GLB</b>	0.59	<b>Asymptotic GLB coef</b>	0.59
<b>Items used in GLB proc</b>	5		
<b>Cut-off score</b>	-0.5	<b>Percentage failing</b>	27.07

Tabel 4.22 *Beskrywende statistiek van Opsomming: Weergawe 1 soos bepaal deur TiaPlus (Eerste loodstoetsing)*

<b>Number of persons in test</b>	193	<b>Number of selected items</b>	6
<b>Minimum test score</b>	0	<b>Maximum test score</b>	6
<b>Average test score</b>	2.66	<b>Standard deviation</b>	1.44
<b>Average P-value</b>	44.39	<b>Std. Error of Measurement</b>	1.05
<b>Average Rit</b>	0.52		
<b>Coefficient Alpha</b>	0.47	<b>SE Coeff. Alpha</b>	0.06
<b>GLB</b>	0.55	<b>Asymptotic GLB coef</b>	0.53
<b>Items used in GLB proc</b>	6		
<b>Cut-off score</b>	0.5	<b>Percentage failing</b>	4.66

Tabel 4.23 *Beskrywende statistiek van Opsomming: Weergawe 2 soos bepaal deur TiaPlus (Eerste loodstoetsing)*

Weergawe 2 se “item fit”-statistiek (kyk tabel 4.24) soos bereken deur Rasch-analise dien as verdere bevestiging van hierdie subtoets se metingspotensiaal. Al die items se “Infit MNSQ”-waardes val binne die voorkeurparameters van 0.5 en 1.5 wat aanduidend is van geskikte passing ten opsigte van die verwagte responspatroon. Die logit-metings van die items wissel van -2.24 (Item 1) tot 1.43 (Item 3) en die metingsfoutwaarde van 0.17 tot 0.21 (“Model SE”). Hierdie foutwaardes is aanvaarbaar, aangesien die foutgrens vir produktiewe items lê tussen -0.3 en 0.3 (Green, 2013:169). Omdat daar egter ’n gaping van -2.01 logits tussen die maklikste item (Item 1) en die volgende item (Item 4) is, kan Item 1 as ’n potensieel problematiese item beskou word, aangesien dit, in vergelyking met die res, die minste konvergent optree. Hierdie item sou dus vir hersiening of weglating oorweeg kon word.

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	PTMEASUR-CORR.	AL-EXP.	EXACT OBS%	MATCH EXP%	ITEM	
6	47	193	1.27	.20	1.07	.7	1.22	1.1	A	.44	.49	81.0	79.7	6
3	43	193	1.43	.20	1.09	.9	1.08	.4	B	.44	.48	78.8	81.3	3
4	93	193	-.23	.17	1.02	.3	.97	-.2	C	.54	.54	66.5	72.0	4
5	89	193	-.11	.17	1.00	.0	1.01	.2	c	.54	.54	74.3	71.6	5
2	90	193	-.14	.17	.93	-1.0	.95	-.4	b	.58	.54	74.9	71.7	2
1	152	193	-2.24	.21	.91	-.7	.62	-1.5	a	.57	.50	82.1	82.1	1
MEAN	85.7	193.0	.00	.19	1.00	.0	.98	-.1				76.3	76.4	
P.SD	36.1	.0	1.21	.02	.07	.7	.18	.8				5.2	4.7	

Tabel 4.24 *“Item fit”-statistiek vir die items van Opsomming, Weergawe 2 soos bepaal deur Winsteps (Eerste loodstoetsing)*

### 3.8 Verwysings

In ooreenstemming met die standaard wat tot dusver gehandhaaf is, lewer hierdie taaktipe eweneens statistiese resultate van bevredigende gehalte op. Die items is aanvanklik in twee afsonderlike helftes geloods op 'n populasie van 129 studente, en hulle produktiwiteit blyk grootliks ekwivalent te wees, aangesien dit oor byna identiese gemiddelde fasiliteits- en diskriminasiewaardes beskik (kyk tabel 4.25). Die gemiddelde r-pbis-waarde is in beide gevalle hoër as 0.5 met 'n gesamentlike gemiddeld van 0.51.

Subtoets	Items	Gem.	SA	Min. Telling	Maks. Telling	Gem. P	Gem. Rpbis
Items in totaal	20	8.97	5.53	1	20	0.45	0.51
Verwysing (1)	10	4.64	2.93	0	10	0.46	0.51
Verwysing (2)	10	4.33	2.96	0	10	0.43	0.52

Tabel 4.25 *Beskrywende statistiek van die Verwysing-subtoets soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)*

Daarbenewens is geen items in enige van die twee assesserings gedurende die analiseproses as probleemitems geïdentifiseer nie, 'n resultaat wat nie anders kan as om 'n bemerkbaar positiewe invloed op die betroubaarheid van die subtoets te hê nie, soos duidelik uit die inligting in tabel 4.26 blyk. Afsonderlik het beide die assesserings Alpha-waardes wat bokant die 0.8-kerf lê, terwyl die gesamentlike Alpha-koëffisiënt grens aan 0.9. Bevredigende resultate soos hierdie maak uiteindelik die deur oop na 'n moeitelose seleksieproses van geskikte items vir 'n verfynde weergawe van hierdie subtoets.

Subtoets	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Rando m	S-B First-Last	S-B Odd-Even
Items in totaal	0.89	1.80	0.82	0.76	0.79	0.90	0.87	0.88
Verwysing (1)	0.81	1.28	0.80	0.53	0.72	0.89	0.70	0.84
Verwysing (2)	0.82	1.25	0.73	0.47	0.76	0.84	0.64	0.86

Tabel 4.26 *Betroubaarheidskoëffisiënte van die Verwysing-subtoets soos bepaal deur Iteman 4.3 (Eerste loodstoetsing)*

#### 4. Die verfyningsproses: 'n Seleksie van produktiewe toetsitems

In die gegewe geval het die verfyningsproses merendeels neergekom op 'n seleksie- eerder as 'n hersieningsproses. Die vernaamste rede hiervoor is dat die eerste loodstoetsing 'n voldoende getal produktiewe items opgelewer het wat boonop in toereikende mate verteenwoordigend blyk te wees van die onderskeie komponente van die toetskonstruk. Die moontlikheid word egter nie uitgesluit dat die geselekteerde items ná verdere loodsing wel toekomstige aanpassings mag ondergaan nie.

Die seleksieproses is uitgevoer ooreenkomstig die kriteria van Van Dyk en Weideman (2004b:17-18) wat in afdeling 2.4 van hierdie hoofstuk bespreek is. Kortom is kwalitatiewe en kwantitatiewe oorwegings in aanmerking geneem wat insluit potensiele items se belyning met die toetskonstruk en -spesifikasies en hulle produktiwiteitsvlak wat bepaal kan word op grond van óf toetsontwerpers se ondervinding en verwagtinge (kwalitatief), óf kwantitatiewe data oor die items, óf 'n kombinasie van beide.

Die volgende tabel gee 'n opsomming van die belangrikste itemstatistiek van die items wat na afloop van die eerste loodstoetsing vir insluiting in die verfynde weergawe van TAGNaS geselekteer is:

	Subtoets	Item ID	Totale Rpbis / Rit	P	Uitgewys / 'Flagged' (J/N)
Toetsfase 1	Skommeltekste	1	0.52 / 0.69	0.66	N
		2	0.70 / 0.82	0.57	N
		3	0.71 / 0.83	0.53	N
		4	0.69 / 0.81	0.44	N
		5	0.59 / 0.75	0.44	N
		6	0.38 / 0.58	0.76	N
		7	0.57 / 0.74	0.57	N
		8	0.69 / 0.82	0.43	N
		9	0.62 / 0.77	0.37	N

		10	0.62 / 0.77	0.36	N	
	<b>Akademiese woordeskat</b>	11	0.32 / 0.41	0.35	N	
		12	0.28 / 0.37	0.58	N	
		13	0.35 / 0.44	0.46	N	
		14	0.43 / 0.51	0.57	N	
		15	0.15 / 0.24	0.71	N	
		16	0.25 / 0.35	0.54	N	
		17	0.17 / 0.27	0.63	N	
		18	0.23 / 0.33	0.53	N	
		19	0.40 / 0.52	0.65	N	
		20	0.21 / 0.35	0.33	N	
	<b>Grammatika en teksverband</b>	21	0.46 / 0.47	0.76	N	
		22	0.48 / 0.51	0.78	N	
		23	0.16 / 0.19	0.34	N	
		24	0.22 / 0.29	0.57	N	
		25	0.33 / 0.42	0.81	N	
		26	0.29 / 0.35	0.60	N	
		27	0.43 / 0.48	0.57	N	
		28	0.32 / 0.37	0.52	N	
		29	0.44 / 0.44	0.65	N	
		30	0.43 / 0.45	0.69	N	
		31	0.47 / 0.50	0.83	N	
		32	0.11 / 0.12	0.42	J	
		33	0.18 / 0.15	0.30	N	
		34	0.37 / 0.35	0.33	N	
		36	0.59 / 0.58	0.77	N	
		36	0.26 / 0.31	0.44	N	
		37	0.47 / 0.50	0.63	N	
		38	0.38 / 0.46	0.67	N	
	<b>Toetsfase 2</b>	<b>Skommelteks</b>	1	0.48 / 0.67	0.66	N
			2	0.63 / 0.78	0.46	N
			3	0.57 / 0.74	0.41	N
			4	0.68 / 0.80	0.33	N
			5	0.63 / 0.77	0.35	N
		<b>Interpretasie van visuele inligting</b>	6	0.45 / 0.52	0.60	N
			7	0.41 / 0.48	0.42	N
			8	0.53 / 0.59	0.59	N
			9	0.34 / 0.41	0.55	N
10			0.45 / 0.51	0.28	N	
11			0.41 / 0.48	0.46	N	



		12	0.49 / 0.56	0.51	N
		13	0.17 / 0.24	0.28	N
		14	0.34 / 0.42	0.54	N
		15	0.20 / 0.28	0.33	N
		16	0.42 / 0.49	0.43	N
		17	0.34 / 0.41	0.54	N
	<b>Akademiese woordeskat</b>	18	0.24 / 0.34	0.52	N
		19	0.30 / 0.39	0.27	N
		20	0.19 / 0.29	0.61	N
		21	0.28 / 0.38	0.53	N
		22	0.35 / 0.44	0.60	N
		23	0.20 / 0.30	0.57	N
		24	0.29 / 0.37	0.30	N
		25	0.26 / 0.35	0.57	N
		26	0.21 / 0.32	0.16	N
		27	0.24 / 0.38	0.51	N
	<b>Tekstipe</b>	28	0.56 / 0.77	0.38	N
		29	0.37 / 0.64	0.65	N
		30	0.43 / 0.70	0.50	N
		31	0.50 / 0.72	0.69	N
		32	0.47 / 0.68	0.75	N
	<b>Teksbegrip</b>	33	0.37 / 0.46	0.54	N
		34	0.27 / 0.35	0.66	N
		35	0.29 / 0.38	0.67	N
		36	0.20 / 0.29	0.58	N
		37	0.29 / 0.37	0.72	N
		38	0.29 / 0.38	0.59	N
		39	0.42 / 0.50	0.44	N
		40	0.36 / 0.46	0.47	N
		41	0.24 / 0.32	0.24	N
		42	0.33 / 0.39	0.83	N
		43	0.32 / 0.39	0.77	N
		44	0.23 / 0.32	0.54	N
		45	0.40 / 0.48	0.38	N
		46	0.41 / 0.49	0.77	N
		47	0.30 / 0.39	0.55	N
		48	0.37 / 0.45	0.42	N
		49	0.39 / 0.47	0.48	N
		50	0.38 / 0.44	0.21	N
		51	0.23 / 0.32	0.75	N

		52	0.23 / 0.32	0.29	N
		53	0.27 / 0.35	0.29	N
		54	0.25 / 0.34	0.64	N
		55	0.15 / 0.25	0.56	N
		56	0.35 / 0.45	0.39	N
		57	0.26 / 0.35	0.45	N
		58	0.31 / 0.41	0.42	N
		59	0.37 / 0.46	0.52	N
		60	0.38 / 0.46	0.71	N
		61	0.40 / 0.49	0.51	N
		62	0.42 / 0.50	0.46	N
	<b>Vergelyking tussen tekste</b>	63	0.16 / 0.25	0.74	N
		64	0.25 / 0.35	0.38	N
		65	0.19 / 0.29	0.33	N
		66	0.50 / 0.56	0.39	N
		67	0.24 / 0.33	0.41	N
	<b>Grammatika en teksverband</b>	68	0.21 / 0.24	0.35	N
		69	0.22 / 0.25	0.35	N
		70	0.38 / 0.59	0.72	N
		71	0.40 / 0.45	0.65	N
		72	0.45 / 0.51	0.63	N
		73	0.54 / 0.56	0.57	N
		74	0.49 / 0.50	0.78	N
		75	0.19 / 0.20	0.39	N
		76	0.39 / 0.42	0.60	N
		77	0.18 / 0.27	0.45	N
		78	0.47 / 0.49	0.76	N
		79	0.49 / 0.50	0.71	N
		80	0.56 / 0.57	0.75	N
		81	0.57 / 0.56	0.72	N
		82	0.55 / 0.56	0.73	N
		83	0.52 / 0.50	0.75	N
	<b>Opsommingtaak</b>	84	0.29 / 0.59	0.47	N
		85	0.17 / 0.44	0.22	N
		86	0.23 / 0.54	0.48	N
		87	0.24 / 0.55	0.46	N
		88	0.16 / 0.44	0.24	N

	<b>Verwysings</b>	89	0.51 / 0.57	0.64	N
		90	0.68 / 0.72	0.40	N
		91	0.68 / 0.72	0.38	N
		92	0.59 / 0.64	0.54	N
		93	0.68 / 0.73	0.51	N

*Tabel 4.27 Itemstatistiek van die geselekteerde items vir die verfynde toetsweergawe van TAGNaS*

Slegs een item uit die bostaande seleksie van items se statistiek val buite die geldende parameters van hierdie studie en het derhalwe 'n sekere mate van herskrywing ondergaan (Item 32, Toetsfase 1). Om die diskriminasiewaarde van hierdie item te probeer verbeter, is een van die afleiers met 'n meer geskikte opsie vervang.

## **5. Slot**

Die doel van hierdie hoofstuk was om 'n oorsig te bied van die resultate van die eerste loodstoetsing van die onderskeie taaktipes wat vir TAGNaS ontwerp is. Aan die hand van statistiese data is sekere gevolgtrekkings omtrent die produktiwiteit van die verskillende subtoetse en hulle individuele items gemaak. Op grond hiervan is daar vir die verfynde, meerfasige weergawe van die toets 'n seleksie van produktiewe toetsitems gemaak.

Ondanks die bevredigende resultate wat behaal is, kan die verfyningsproses egter in hierdie stadium geensins as volkome afgehandel beskou word nie. Die verfynde toetsweergawe, of ten minste 'n aanvaarbare gedeelte daarvan, moet verkieslik aan herloodsing ('repiloting') onderwerp word. Dit is die betroubaarste manier waarop vasgestel kan word of die geselekteerde items voldoende antwoorde kan verskaf op die vrae wat opgewerp is deur die aanvanklike statistiese analises. Ná herloodsing sal die analiseproses hervat word en, indien nodig, sal verdere aanpassings aan die individuele items gemaak word. Daar word spesifiek in die volgende hoofstuk van hierdie studie verslag gedoen van die herloodsing van die verfynde toetsweergawe van TAGNaS, asook van

enige verdere verfynings aan die ontwerp, waarna aangedui sal word of die toets, soos dit tans daar uitsien, gereed is vir die werklike administrasiefase al dan nie.

## **Hoofstuk 5**

### **Die verdere verfyning van 'n nagraadse toets van akademiese geletterdheid in Afrikaans: Resultate van 'n tweede loodstoets**

#### **1. Inleiding**

In die vorige hoofstuk is die statistiese resultate van die eerste loodstoetsing van die Toets van Akademiese Geletterdheid vir Nagraadse Studente (TAGNaS) gerapporteer, saam met besonderhede oor die daaropvolgende verfynings-/seleksieproses wat plaasgevind het. Nog 'n rondte van loodstoetsing, hierdie keer van die verfynde weergawe van die toets, is ook onderneem waarvan die resultate in hierdie hoofstuk uiteengesit word. Green (2014:44) beklemtoon dat “[t]rials are vital to the quality of any kind of assessment and should always be carried out to the extent that is possible”. Trouens, Green (2014:44) lys verskeie voorstelle met betrekking tot die loodsing en verfyning van taaltoetse waaraan daar in hierdie studie voldoen word. Hy stel onder andere voor dat aansienlik meer materiaal geloods moet word as wat uiteindelik benodig gaan word om sodoende die kanse op die insluiting van genoeg produktiewe items in 'n bepaalde assessering te vergroot.

Vir die verfynde toets in hierdie studie is 131 items van die 385 items wat oorspronklik geloods is vir verdere loodsing geselekteer; persentasiegewys beteken dit dat 66% meer items ontwerp is as wat uiteindelik by die tweede, verfynde loodstoets ingesluit is. Verder, in ooreenstemming met 'n bykomende voorstel van Green (2014:44), is slegs produktiewe items vir die verfynde weergawe geselekteer, terwyl minder produktiewe items weggelaat of in hersiening geneem is. Wat betref die loodsing van 'n verfynde toets, vereis Green (2014:144) pertinent “always pilot the material again before using it in an operational assessment”.

Hoofstuk 5 bied, in ooreenstemming met die metodologie van die vorige hoofstuk, 'n oorsig van die resultate van die loodstoetsing van die verfynde weergawe van TAGNaS. Dieselfde benadering word gevolg deurdat daar weereens aan die hand van statistiese data sekere gevolgtrekkings omtrent die produktiwiteit en betroubaarheid van die toets in sy geheel en van die onderskeie subtoetse afsonderlik gemaak word. Omdat TAGNaS as 'n potensiële meerfasige toets beplan word, is dit van wesenlike belang geag om die verfynde toets in hierdie vorm te loods. Gevolglik word die statistiese analyses van die eerste twee toetsdele of -fases (beide ten volle veelkeuseformaat) apart van mekaar aangebied, oorweeg en bespreek.

Let op hoedat daar verskeie kere in die loop van die hieropvolgende bespreking pertinent kommentaar gelewer word op die produktiwiteit al dan nie van die gewysigde en nuwe subtoetse/taaktipes<sup>15</sup> wat in TAGNaS op die proef gestel word. Dit is belangrik om van laasgenoemde kennis te neem, omdat eksperimentering met vernuwende elemente in die operasionalisering van die toetskonstruk een van die hoofdoelwitte van hierdie studie is, soos in Hoofstuk 1 uiteengesit.

## **2. 'n Volgende stap in die toetsontwikkelingsproses: Die tweede loodstoetsing van TAGNaS**

### **2.1 Getalle en deelnemers**

Die verfynde eerste en tweede toetsdele van TAGNaS is gedurende Januarie/Februarie van 2017 op 'n gemengde groep Afrikaanse voorgraadse studente (wat wissel van eerste- tot derdejaarsvlak) aan twee plaaslike tersiêre instansies, geloods, naamlik Universiteit Stellenbosch (US) en Aros, 'n geakkrediteerde Afrikaanse private hoëronderrwysinstelling vir

---

15. Hierdie taaktipes betrek veral analitiese denke ten einde met die analitiese karakter van akademiese diskoers, wat aan die kern van die toetskonstruk lê, te bely.

onderwysstudente in Pretoria. Die eerste gedeelte van die verfynde toets wat as 'n verpligte of opsionele siftingstoets oorweeg word<sup>16</sup>, is deur 'n kohort van 271 kandidate afgelê, en die tweede, langer, meer deeglike assessering deur byna 100 kandidate. Die totale kohort vir die loodsing van die verfynde toets het bestaan uit 370 kandidate. Hoewel hierdie kohort redelik groot van omvang is, word die aanbeveling gemaak dat verdere herloodsing van die verfynde toets sover moontlik op nog geskikte kandidate uitgevoer word soos wat loodseleenthede hulle in die toekoms sou voordoën. Vir die doeleindes van hierdie studie word die grootte van hierdie kohort egter as bevredigend geag, aangesien die statistiese resultate wat opgelewer is van genoeg betekenis is om voorlopig geldige gevolgtrekkings omtrent die effektiwiteit en betroubaarheid van die verfynde toets te maak.

## 2.2 Procedure, benadering en parameters

Beide toetsdele is in geheel in een sitting op twee afsonderlike geleenthede deur verskillende toetsgroepe voltooi. Optiese leeskaarte wat elektronies by die Universiteit van die Vrystaat (UV) nagesien word, is weereens as antwoordstelsel gebruik, en die toetsresultate vir statistiese analise in MS Excel-formaat uitgebring. Die data-analise is deur Iteman 4.3 en Tiaplus uitgevoer; dus vanuit die benadering van klassieke toetsteorie (KTT). Die grondbeginsels van KTT is reeds in detail in die vorige hoofstuk uiteengesit (kyk afdeling 2.3), asook die psigometriese eienskappe op grond waarvan toetsitems se produktiwiteit in hierdie benadering bepaal word (kyk afdeling 2.4). In kort berus die analise van die produktiwiteit van die loodstoetsitems in hierdie studie primêr op die volgende vier vrae en die parameters wat vir elkeen geld:

- Is die items geskik in terme van hulle fasiliteitswaarde? (P-waarde tussen 0.15 en 0.84)

---

16. Kyk weer na tabel 3.4 in Hoofstuk 3 vir 'n uiteensetting van TAGNaS as 'n potensiële driefasetoets.

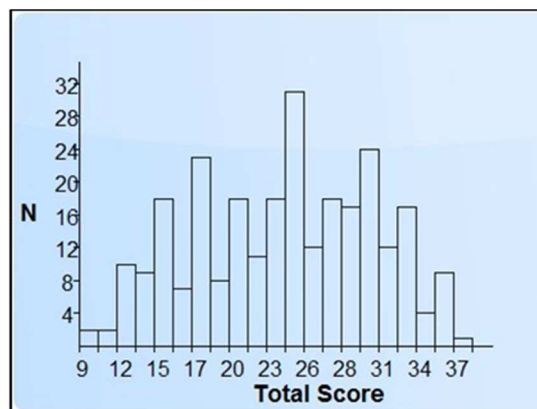
- Is die diskriminasiewaarde van die items voldoende/bevredigend? (minimum r-pbis-waarde van 0.15, selde hoër as 0.50)
- Hoe goed is die betroubaarheidskoeffisiënt van elke subtoets/die toets in sy geheel? (Cronbach se Alpha en ‘Greatest Lower Bound’, GLB)
- Is die interne korrelasies (bv. tussen die items/subtoetse onderling) bevredigend?

### 3. Die tweede loodstoetsing van TAGNaS: Analise en bespreking van toetsresultate

Vanweë die meerfasige formaat van die verfynde toets word die statistiese loodsdata van die eerste twee toetsfases afsonderlik van mekaar aangebied en bespreek.

#### 3.1 Statistiese resultate van eerste toetsfase (verpligte/opsionele siftingstoets)

Figuur 5.1 toon die verspreiding van die puntetellings van die kandidate vir die totale aantal items oor al die subtoetse heen.



*Figuur 5.1 Frekwensiedistribusie van totale toetstellings (Toetsfase 1) soos bepaal deur Iteman 4.3*

Die gemiddelde totale toetstelling word aangegee as 24.93, terwyl die meeste kandidate 'n totale telling van tussen 25 en 26 vir die toets behaal het. Soos uit die kolomgrafiek waargeneem kan word, is die verspreidingskromme redelik



gespits (kurtose) en beskik dit oor negatiewe skeefheid (d.i. 'n stert na links). Volgens TiaPlus is die kurtosewaarde van hierdie toets gelyk aan 2.14 en die skeefheidswaarde word aangedui as -0.17. Hoewel selfs die beste taaltoets noodwendig 'n sekere mate van skeefheid in sy distribusie sal vertoon, meen Bachman (2004:74) dat, volgens die praktyk, “values for skewness and kurtosis of between -2 and +2 indicate a reasonably normal distribution”. Gemeet aan hierdie parameters, beskik hierdie toets oor 'n redelik normale verspreiding, hoewel die aangeduide kurtosewaarde effens hoër as die preferente positiewe perk is. Hierdie afwyking behoort nie geïgnoreer te word indien dit meer sou vergroot soos wat verdere eksperimentering met die siftingstoets plaasvind nie, aangesien 'n groot negatiewe skeefheid daarop kan dui dat 'n toets te veel maklike items bevat. Die gemiddelde P-waarde van die toets wat op 0.66 vasgestel word (kyk tabel 5.1), besweer egter vir eers hierdie besorgdheid omdat hierdie waarde nie werklik as uitermatig hoog bestempel kan word nie. Omdat hierdie toets in alle geval as 'n siftingstoets beoog word, geld die aanbeveling buitendien dat die afsnypunt daarvan redelik hoog gestel moet word sodat dit hoofsaaklik die min-tot-geen-risiko-studente is wat reeds vroeg daardeur geïdentifiseer en uit die toetsingsproses uitgeskakel word. Daar is reeds oor hierdie aspek uitgebrei in Hoofstuk 3. Alternatiewelik kan 'n laer gemiddelde P-waarde ook verkry word deur die puntetoekenning van sekere moeiliker subtoetse of individuele items opwaarts aan te pas, of bloot deur die makliker subtoetse (bv. moontlik Skommelteks 1) of individuele items verder te verfyn of te hersien.

Subtoets	Items	Gem.	SA	Min. telling	Maks. telling	Gem. P	Gem. Rpbis
Alle items	38	24.93	6.92	9	38	0.66	0.35
Skommelteks 1	5	3.56	1.77	0	5	0.71	0.44
Skommelteks 2	5	3.21	1.74	0	5	0.64	0.42
Woordeskat (Enkele gaping)	8	4.93	1.76	1	8	0.62	0.25
Woordeskat (Dubbele gaping)	2	1.14	0.65	0	2	0.57	0.29
Grammatika & teksverband	18	12.08	3.88	0	18	0.67	0.35

*Tabel 5.1 Beskrywende statistiek vir Toetsfase 1 soos bepaal deur Iteman 4.3*

Die gemiddelde diskriminasiewaarde van die toets ('n r-pbis van 0.35), en van die onderskeie subtoetse, is nietemin ook hoogs bevredigend en almal bo die geldende minimumparameter van 0.15. TiaPlus dui die gemiddelde Rit-waarde van die toets in sy geheel effens hoër aan as 0.40. Wat veral opvallend is uit die data hierbo, is die r-pbis-waarde van 0.29 van die 'dubbele gaping'-woordeskatsubtoets – 'n gewysigde taaktipe wat vir die eerste keer in TAGNaS op die proef gestel word – veral in die lig daarvan dat hierdie subtoets uit slegs twee items bestaan, en dat die diskriminasiewaarde van subtoetse of items oor die algemeen selde hoër as 0.5 r-pbis uitreken (Guyer & Thompson, 2011:30). Hierdie dien as nog 'n bewys van die toenemend klaarblyklike slotsom dat eksperimentering met dié aangepaste itemtipe wel as geregverdig beskou kan word. Hierdie gevolgtrekking hou steek, selfs in weerwil van die laer Alpha-waarde van hierdie subtoets soos blyk uit tabel 5.2. 'n Langer (sub)toets is oor die algemeen meer betroubaar as 'n korter een, en omdat hierdie subtoets uit slegs twee items bestaan, is 'n laer betroubaarheidswaarde hier noodwendig te wagte. Dieselfde met die 'enkelgaping'-woordeskatsubtoets wat uit slegs agt items bestaan. Dit sou dus onverstandig wees om hierdie twee subtoetse voor die voet, en uitsluitlik op grond van hulle laer betroubaarheidswaardes te verwerp as onbetroubare indikatore van kandidate se akademiese woordeskatvermoë. Bowendien, indien die twee woordeskatsubtoetse elk uit 40 items sou bestaan, stel die Spearman-Brown-korreksieformule die Alpha-koëffisiënt daarvan vas op onderskeidelik 0.86 (dubbele gaping) en 0.81 (enkelgaping). Beide hierdie waardes is vir subtoetse meer as aanvaarbaar.

Daar is reeds in Hoofstuk 3 vooruit gewys op die hoë waarskynlikheid van 'n tegnies gekwalifiseerde wisselspel tussen die kinematiese en ekonomiese ontwerpaspekte by die samestelling van 'n korter siftingstoets – 'n omstandigheid wat hom nou kennelik hier aan die afspeel is. Daar is vroeër daarop gewys dat, in die geval van korter siftingstoetse, ontwikkelaars hoogs waarskynlik met laer (maar steeds aanvaarbare) betroubaarheidswaardes tevrede sal moet wees – 'n kompromie wat in die betrokke geval nie werklikwaar die toets se status as regverdige meetinstrument blyk te ondermyn nie, aangesien dit steeds oor 'n hoogs bevredigende algehele Alpha-waarde van 0.86 beskik. Trouens, TiaPlus stel die algehele GLB-koëffisiënt nóg hoër vas op 0.95. Die oorblywende subtoetse (skommeltekste, grammatika en teksverband) se individuele Alpha-waardes is ook meer as voldoende, aangesien almal bokant 0.8 lê.

Subtoets	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Alle items	0.86	2.60	0.69	0.45	0.85	0.82	0.62	0.92
Skommelteks 1	0.85	0.68	0.81	0.59	0.81	0.89	0.75	0.89
Skommelteks 2	0.83	0.72	0.52	0.52	0.83	0.68	0.68	0.91
Woordeskat (Enkele gaping)	0.46	1.29	0.27	0.32	0.33	0.42	0.49	0.49
Woordeskat (Dubbele gaping)	0.25	0.57	0.14	0.14	0.14	0.24	0.24	0.24
Grammatika & teksverband	0.81	1.70	0.66	0.52	0.83	0.79	0.69	0.91

*Tabel 5.2 Betroubaarheidskoëffisiënte van die subtoetse in Toetsfase 1 soos bepaal deur Iteman 4.3*

Wat betref die ooreenkoms van die samestellende subtoetse met mekaar in terme van dit wat gemeet word, bereken Iteman 4.3 die korrelasiekoëffisiënte tussen die verskillende subtoetse soos volg:

Subtoetse	Skommelteks	Skommelteks	Woordeskat	Woorde-	Grammatika &
	1	2	(enkel)	skat	teksverband
				(dubbel)	
Skommelteks 1					
Skommelteks 2	0.41				
Woordeskat (enkel)	0.31	0.35			
Woordeskat (dubbel)	0.21	<b>0.17</b>	0.28		
Grammatika & teksverband	0.29	0.28	0.41	0.34	

*Tabel 5.3 Interne korrelasies tussen die subtoetse van Toetsfase 1 soos bepaal deur Iteman 4.3*

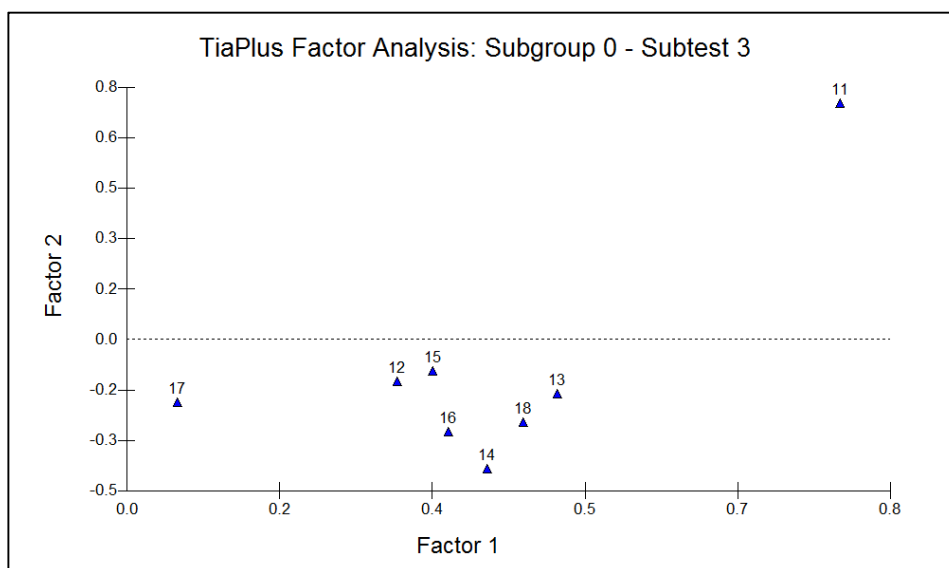
Met die uitsondering van een interkorrelasie (tussen die tweede skommelteks en die ‘dubbele gaping’-woordeskatsubtoets), lê al die subtoets-interkorrelasies in tabel 5.3 tussen die voorkeurparameters van 0.2 en 0.5. Van der Walt en Steyn (2007:148) (kyk ook Steyn, 2014:33) voer aan dat, indien subtoets-interkorrelasies aan hierdie parameters voldoen, dit daarop dui dat elke subtoets in ’n toets ’n ander aspek van die konstruk meet – ’n afleiding wat berus op die uitgangspunt van Alderson et al. (1995:184) dat “since the reason for having different test components is that they all measure something different and therefore contribute to the overall picture of language ability attempted by the test, we should expect these correlations to be fairly low”. Tussen die subtoetse onderling lyk dit dus of hierdie toets ’n geskikte mate van heterogeniteit of geskakeerdheid vertoon.

Subtoetse	Subtoets	Toets as geheel	Subtoetse				
			1	2	3	4	5
Skommelteks 1	1	0.58					
Skommelteks 2	2	0.58	0.41				
Woordeskat (enkel)	3	0.69	0.31	0.35			
Woordeskat (dubbel)	4	0.46	0.22	0.17	0.28		
Grammatika & teksverband	5	0.87	0.29	0.28	0.41	0.34	

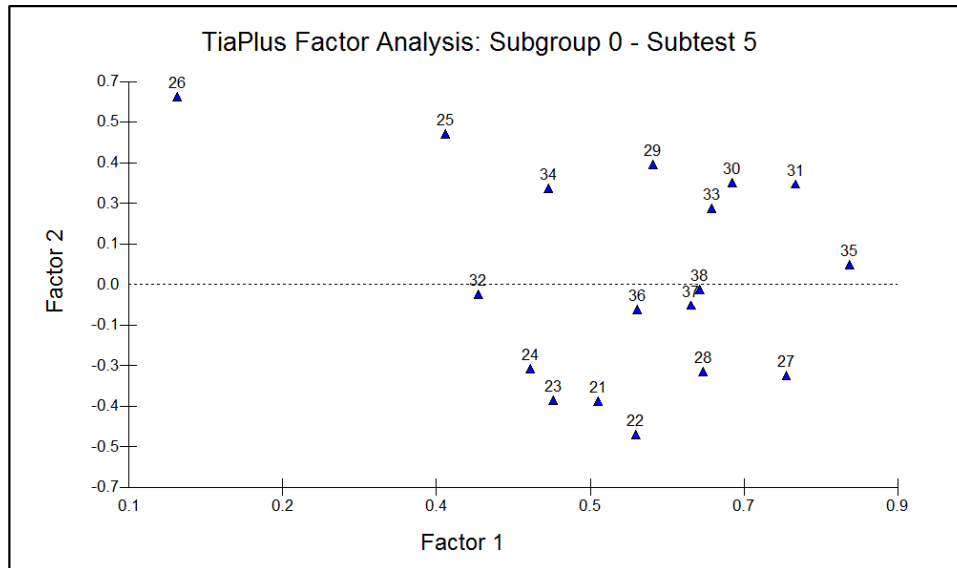
*Tabel 5.4 Interne korrelasies tussen die subtoetse van Toetsfase 1 en die toets as geheel soos bepaal deur TiaPlus*

Die inligting in tabel 5.4, saam met faktoranalises van die onderskeie subtoetse self, bevestig egter tegelykertyd ook dat die toets redelik homogeen is in terme

van dit wat gemeet word. Die korrelasies tussen die verskillende subtoetse en die toets as geheel stem besonder goed ooreen met die norm van 0.6 of hoër, met slegs een uitsondering (woordeskate dubbel), en om verklaarbare redes (klein aantal items). Figuur 5.2 en 5.3 hieronder vertoon die spreidingsdiagramme vir die faktoranalises van die woordeskate- (enkelgaping) en grammatika en teksverband-subtoetse wat deur TiaPlus gegenereer is. Behalwe vir Items 11 en 26 wat weg van die groter groeperings uitstip, kluster die res van die items in elk van hierdie subtoetse grotendeels in dieselfde domein. Hoewel uitskieters soos hierdie ook die resultaat kan wees van die heterogeniteit van die toetskonstruk, word daar nietemin aanbeveel dat hierdie twee items, ná verdere herloodsing van die eerste toetsfase, wéér krities ondersoek moet word ten einde te bepaal of die items verdere verfyning benodig al dan nie.



*Figuur 5.2 Faktoranalise: Woordeskate – Enkelgaping (Tweede loodstoetsing)*



*Figuur 5.3 Faktoranalise: Grammatika en teksverband (Tweede loodstoetsing)*

'n Laaste belangrike statistiese eienskap van die loodstoets wat in die onderhawige toetsfase oorweeg moet word, is die omvang van sy misklassifikasies. Omdat geen toets volkome betroubaar is nie is 'n belangrike potensiële negatiewe gevolg van die administrasie daarvan dat dit die kandidate wat dit aflê verkeerdelik kan klassifiseer. Van der Slik en Weideman (2005:28) verduidelik dat “because its reliability is never 100%, a test may misclassify a failure as a pass, and vice versa”. Tabel 5.5 wys die aantal potensiële misklassifikasies soos bereken vir die eerste toetsfase gebaseer op vier verskillende kriteria.

<b>Alpha-gebaseer:</b>	
Korrelasie tussen toets en hipotetiese, parallelle toets	10 (3.8%)
Korrelasie tussen waargenome en 'werklike' toetstellings	8 (2.8%)
<b>GLB-gebaseer:</b>	
Korrelasie tussen toets en hipotetiese, parallelle toets	7 (2.6%)
Korrelasie tussen waargenome en 'werklike' toetstellings	5 (1.9%)

*Tabel 5.5 Potensiële misklassifikasies van Toetsfase 1 soos bepaal deur TiaPlus*

Soos gesien kan word, wissel die aantal potensiële misklassifikasies (vir die betrokke kohort van 271 studente) tussen 5 en 10, afhangende van watter kriteria

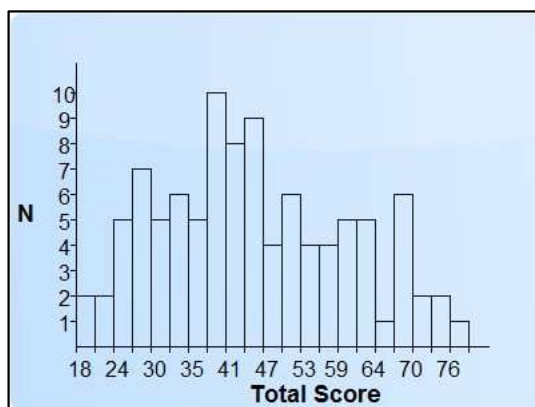
toegepas word. Van der Slik en Weideman (2005:29) herinner egter dat ongeveer die helfte van die potensiële misklassifikasies ontstaan as gevolg van kandidate wat geslaag het, terwyl hulle eintlik moes druip. Dit beteken met ander woorde dat tussen 3 en 5 kandidate wat moes slaag moontlik gedruip het. Wanneer die afsnypunt van die toetsfase dus uiteindelik bepaal word, kan die aantal kandidate wat 'n tweede kans gegun behoort te word, op dieselfde manier bereken word. In die geval van 'n driefasige toets kan dit op meer as een manier hanteer word. Grensgeval-kandidate kan byvoorbeeld toegelaat word om die tweede, langer toetsfase oor te slaan, maar hulle moet dan steeds die derde toetsfase (d.i. die skryftaak) voltooi. Vanuit 'n teoretiese oogpunt is die inagneming van potensiële misklassifikasies 'n tegniese oorweging wat uit die juridiese en etiese analogieë resulteer, d.i. dit hou verband met die ontwerpbeginsels van regverdigheid en sorgsaamheid.

### 3.2 Statistiese resultate van tweede toetsfase

Die loodsingresultate van die tweede toetsfase toon 'n tellingverspreiding waarvan die kromme redelik gespits (kurtose) en effens positief skeefgetrek (d.i. met 'n stert na regs) is. Figuur 5.4 (kyk volgende bladsye) wys die verspreiding van die puntetellings van die kandidate vir die totale aantal items oor al die subtoetse heen. Die gemiddelde totale toetstelling word bereken as 47.03, terwyl die grootste getal studente 'n telling van tussen 40 en 42 vir die toets behaal het. TiaPlus dui die kurtosewaarde van hierdie toets aan as 2.30 en die grootte van die positiewe skeefheid as 0.37. Die toets beskik dus oor 'n redelik normale verspreiding, hoewel die kurtosewaarde wel 'n klein bietjie hoër as die voorkeurmaksimum van +2 is (Bachman, 2004:74), 'n afwyking waarvan die omvang dopgehou behoort te word soos wat verdere loodsing van die verfynde toets op groter studentekohorte plaasvind. Die positiewe skeefheid, hoewel binne aanvaarbare perke, dui daarop dat die toets effens moeilik vir die betrokke kohort was – 'n uitslag wat nie werklik as onverwags beskou kan word nie, aangesien daar by die seleksie van items vir die verfynde toets voorkeur gegee is aan

produktiewe items met laer P-waardes, d.i. meer uitdagende items, gesien in die lig daarvan dat die toets uiteindelik vir seniorvlak studente bedoel is.

Nietemin lê die gemiddelde fasiliteitswaarde van die toets steeds naby aan 0.50 (kyk tabel 5.6) wat gewoonlik die mikpunt van gestandaardiseerde toetse is, aangesien hierdie gemiddeld toelaat vir maksimum differensiasie onder toetskandidate (Davies et al., 1999:95-96). Wat die moeilikheidsgraadverspreiding van die individuele subtoetse betref, wissel die gemiddelde P-waardes van 0.36 tot 0.72, dus almal val binne die geldende statistiese grense van hierdie studie en kom ooreen met die voorkeurverspreiding wat vir gestandaardiseerde taaltoetse geld, naamlik van 0.3 tot 0.7 (Davies et al., 1999:95-96). Dit is wel so dat vir die meerderheid van die subtoetse (sewe uit die tien) die individuele P-waardes onder die algehele gemiddeld van 0.51 lê, wat aanduidend is van subtoetse met 'n hoër moeilikheidsgraad. Hierdie besonderheid strook met die positiewe skeefheid van die punteverspreiding waaroor reeds verhelderende kommentaar hierbo gelewer is.



*Figuur 5.4 Frekwensiedistribusie van totale toetstellings (Toetsfase 2) soos bepaal deur Ite-man 4.3*



Subtoets	Items	Gem. SA	Min. Telling	Maks. Telling	Gem. P	Gem. Rpbis	
Alle items	93	47.03	15.01	18	79	0.51	0.32
Skommelteks	5	2.21	1.94	0	5	0.44	0.34
Interpretasie van visuele inligting	12	4.63	2.40	0	11	0.39	0.22
Woordeskat (enkelgaping)	8	4.07	2.23	0	8	0.51	0.38
Woordeskat (dubbele gaping)	2	0.73	0.73	0	2	0.36	0.53
Tekstipe	5	3.60	1.41	0	5	0.72	0.22
Teksbegrip	30	13.92	5.32	5	29	0.46	0.31
Teksbegrip: Vergelyking tussen tekste	5	2.01	1.40	0	5	0.40	0.39
Grammatika & teksverband	16	11.40	3.56	0	16	0.71	0.36
Opsomming	5	2.19	1.33	0	5	0.44	0.30
Verwysings	5	2.27	1.88	0	5	0.45	0.32

Tabel 5.6 *Beskrywende statistiek vir Toetsfase 2 soos bepaal deur Iteman 4.3*

Die gemiddelde diskriminasiewaardes hierbo, van die toets in sy geheel (0.32 r-pbis) en van die onderskeie subtoetse, wys egter dat hierdie toets steeds goed tussen sterker en swakker kandidate diskrimineer (kyk tabel 5.6). Die r-pbis-waardes van die afsonderlike subtoetse wissel van 0.22 tot 0.53, met ander woorde al die waardes lê bokant die minimumparameter van 0.15. Gelet op die feit dat die r-pbis-korrelasie van 'n goeie item of subtoets gewoonlik hoër, maar selde hoër as 0.50 is (Guyer & Thompson, 2011:30), wil dit lyk of die dubbele gaping-woordeskat-subtoets, met sy r-pbis van 0.53, oor uitstekende diskriminasiepotensiaal beskik. Trouens, hierdie subtoets wat een van die gewysigde subtoetse is waarmee daar in TAGNaS geëksperimenteer word, se diskriminasiewaarde is die hoogste van die groep. Die subtoets met die tweede hoogste diskriminasiewaarde is die 'Vergelyking tussen tekste'-subtoets (0.39 r-pbis) wat ook 'n nuwe subtoets is wat in hierdie studie op die proef gestel word. Resultate soos hierdie regverdig in toenemende mate die insluiting van hierdie vernuwende elemente in die onderhawige toetsontwerp.

'n Toets met konsekwente hoë vlakke van itemdiskriminasie, soos wat hier die geval blyk te wees, is gewoonlik 'n betroubare toets. Tabel 5.7 hieronder bevat 'n betroubaarheidsanalise van die tweede toetsfase in sy geheel en van die onderskeie subtoetse individueel. Die algehele Alpha-waarde van hierdie toets word vasgestel op 'n uiters bevredigende 0.92. Dit beteken dat hierdie studie

geslaag het in die doelwit om 'n Afrikaanse eweknie vir die Test of Academic Literacy for Postgraduate Students (TALPS) te produseer wat net soos sy voorganger oor 'n algehele betroubaarheidskoeffisiënt van bo 0.9 beskik. Die individuele subtoetse met die hoogste betroubaarheidsmetings (bokant 0.7) is die skommelteks, die enkelgaping-woordeskattoefening, teksbegrip, grammatika en teksverband en verwysings, waarvan spesifiek laasgenoemde ook 'n nuwe subtoets is waarvan die toetspotensiaal in hierdie studie ondersoek word.

Die laer Alpha-waardes van veral die dubbele gaping-woordeskatsubtoets, die teksvergelyking-subtoets en die opsommingtaak – al drie nuwe taaktipes wat in TAGNaS beproef word – lyk wel met die eerste oogopslag taamlik twyfelagtig, maar daar moet in gedagte gehou word dat hierdie subtoetse almal uit slegs vyf items of minder bestaan. Sodra die Spearman-Brown-korreksieformule by die betroubaarheidsanalise in berekening gebring word, neem die Alpha-koeffisiënt van elk van hierdie assesserings aansienlik toe. Indien hierdie subtoetse elk uit 40 items sou bestaan, sou die Alpha-waardes daarvan volgens TiaPlus respektiewelik as 0.92, 0.90 en 0.84 aangedui gewees het. Hierdie syfers inaggenome, tesame met veral die hoogs bevredigende algehele Alpha-waarde van die toets, kan daar, binne die bestek van hierdie studie, tot die gevolgtrekking gekom word dat die tweede toetsfase van TAGNaS oor 'n aanvaarbare vlak van interne konsistensie beskik.

Subtoets	Alpha	SEM	Split-Half (Random)	Split-Half (First- Last)	Split-Half (Odd- Even)	S-B Random	S-B First- Last	S-B Odd- Even
Alle items	0.92	4.25	0.80	0.73	0.90	0.89	0.84	0.95
Skommelteks	0.86	0.74	0.85	0.65	0.78	0.92	0.79	0.88
Grafiese & visuele inligting	0.62	1.48	0.42	0.53	0.46	0.59	0.70	0.63
Woordeskat (enkel)	0.70	1.23	0.57	0.51	0.51	0.73	0.68	0.67
Woordeskat (dubbel)	0.40	0.56	0.24	0.24	0.24	0.38	0.38	0.38
Tekstipe	0.66	0.82	0.68	0.82	0.32	0.81	0.90	0.49
Teksbegrip	0.78	2.49	0.70	0.67	0.69	0.82	0.80	0.82
Teksbegrip: Vergelyking tekste	0.53	0.96	0.31	0.29	0.50	0.47	0.44	0.66
Grammatika & teksverband	0.82	1.50	0.74	0.59	0.79	0.85	0.74	0.88
Opsomming	0.40	1.03	0.24	0.28	0.38	0.39	0.44	0.55
Verwysings	0.84	0.75	0.77	0.76	0.78	0.87	0.86	0.88

*Tabel 5.7 Betroubaarkheidskoëffisiënte van die subtoetse in Toetsfase 2 soos bepaal deur Iteman 4.3*

Verder dien die interne korrelasies tussen die subtoetse in hierdie toetsfase, tesame met faktoranalises van die verskillende subtoetse, tot bewys dat hierdie toets in terme van sy dimensionaliteit wel daarin slaag om 'n voldoende mate van balans tussen homo- en heterogeniteit te handhaaf. Wat subtoets-interkorrelasies betref, geld die algemene vereiste hier dat die korrelasies verkieslik betreklik laag moet wees (d.i. onder 0.5) omdat die subtoetse in 'n toets nie veronderstel is om almal dieselfde konstrukskomponente te meet nie (Alderson et al., 1995:184). Met uitsondering van enkele van die interkorrelasies in tabel 5.8 lê die meerderheid daarvan vir hierdie toetsfase onder die perk van 0.5. In hierdie tabel is die problematies lae waardes (onder 0.15) in kursief aangedui, en die wat effe te hoog was (bokant 0.55) in vetdruk. Aan die een kant is dit nie verbasend dat subtoetse soos teksbegrip en teksvergelyking, asook teksbegrip en woordeskat 'n ietwat sterker (maar steeds aanvaarbare) wedersydse verwantskap sal toon nie, aangesien daar in hierdie subtoetse se artikulasie met die toetskonstruk 'n sekere mate van oorvleueling is (vgl. die toetsspesifikasies in bylaag B). Terselfdertyd behoort daar ook ruimte gelaat te word vir die paar interkorrelasies wat onder die voorkeurminimumperk van 0.2

lê, omdat die geskakeerdheid van die toetskonstruk, aan die ander kant weer, onvermydelik met 'n sekere mate van heterogeniteit gepaard sal gaan.

Subtoetse	Skommelteks	Interpretasie van visuele inligting	Woordeskat (enkel)	Woordeskat (dubbel)	Tekstipe	Teksbegrip	Vergelyking tussen tekste	Grammatika & teksverband	Opsomming	Verwysings
Skommelteks										
Interpretasie van visuele inligting	0.28									
Woordeskat (enkel)	0.13	0.41								
Woordeskat (dubbel)	0.22	0.39	0.54							
Tekstipe	0.15	0.10	0.25	0.22						
Teksbegrip	0.38	0.49	<b>0.60</b>	<b>0.58</b>	0.35					
Vergelyking tussen tekste	0.22	0.40	0.51	<b>0.55</b>	0.20	<b>0.66</b>				
Grammatika & teksverband	0.23	0.23	0.46	0.52	0.19	<b>0.55</b>	0.44			
Opsomming	0.22	0.23	0.42	0.38	0.09	0.48	0.42	0.49		
Verwysings	0.16	0.08	0.25	0.30	0.11	0.31	0.15	0.34	0.26	

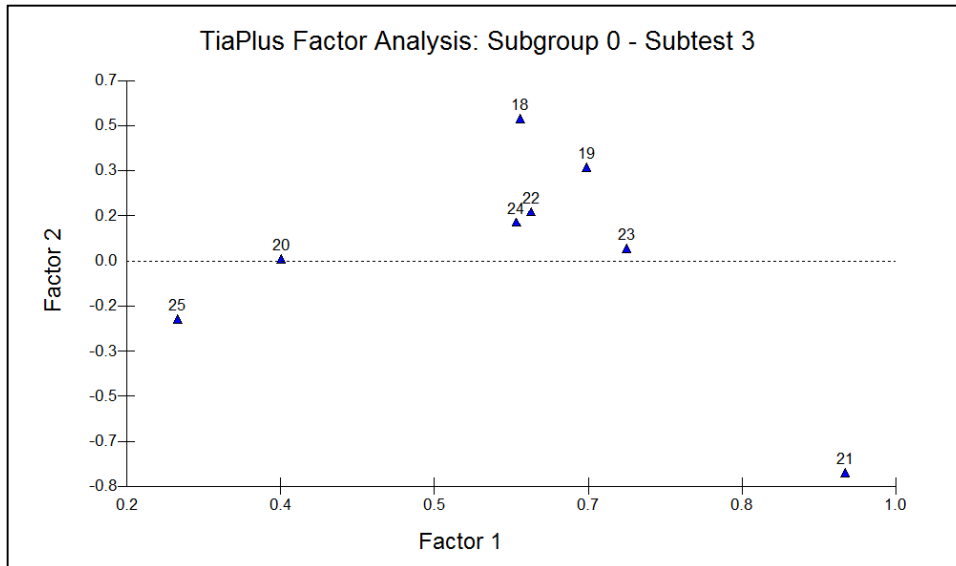
Tabel 5.8 *Interne korrelasies tussen die subtoetse van Toetsfase 2 soos bepaal deur Iteman 4.3*

Die interkorrelasies tussen die verskillende subtoetse en die toets as geheel (kyk tabel 5.9) dui egter ook aan hoedat die toets, oor die geheel geneem, 'n aanvaarbare vlak van dimensionele samehang vertoon. Gemeet aan die norm van 0.6 of hoër is daar slegs drie subtoetse wat nie na behore met die toets as geheel korreleer nie, 'n resultaat wat genoegsaam bevredig, veral in die lig van die grootte van die betrokke kohort.

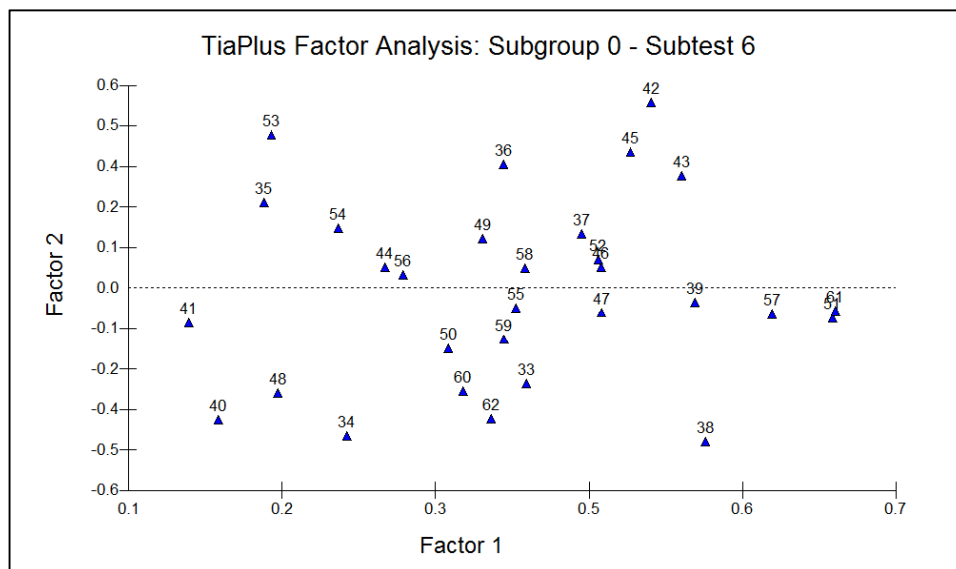
Subtoetse	Subtoets	Toets as geheel	Subtoetse											
			1	2	3	4	5	6	7	8	9	10		
Skommelteks	1	<b>0.44</b>												
Interpretasie van visuele inligting	2	0.59	0.32											
Woordeskat (enkel)	3	0.73	0.13	0.41										
Woordeskat (dubbel)	4	0.72	0.20	0.39	0.54									
Tekstipe	5	<b>0.36</b>	0.11	0.12	0.24	0.22								
Teksbegrip	6	0.90	0.38	0.49	0.60	0.58	0.33							
Teksvergelyking	7	0.75	0.24	0.40	0.51	0.55	0.22	0.66						
Grammatika & teksverband	8	0.73	0.21	0.23	0.46	0.52	0.17	0.55	0.44					
Opsomming	9	0.60	0.21	0.23	0.42	0.38	0.11	0.48	0.42	0.49				
Verwysings	10	<b>0.39</b>	0.12	0.07	0.24	0.29	0.09	0.29	0.15	0.32	0.25			

*Tabel 5.9 Interne korrelasies tussen die subtoetse van Toetsfase 2 en die toets as geheel soos bepaal deur TiaPlus*

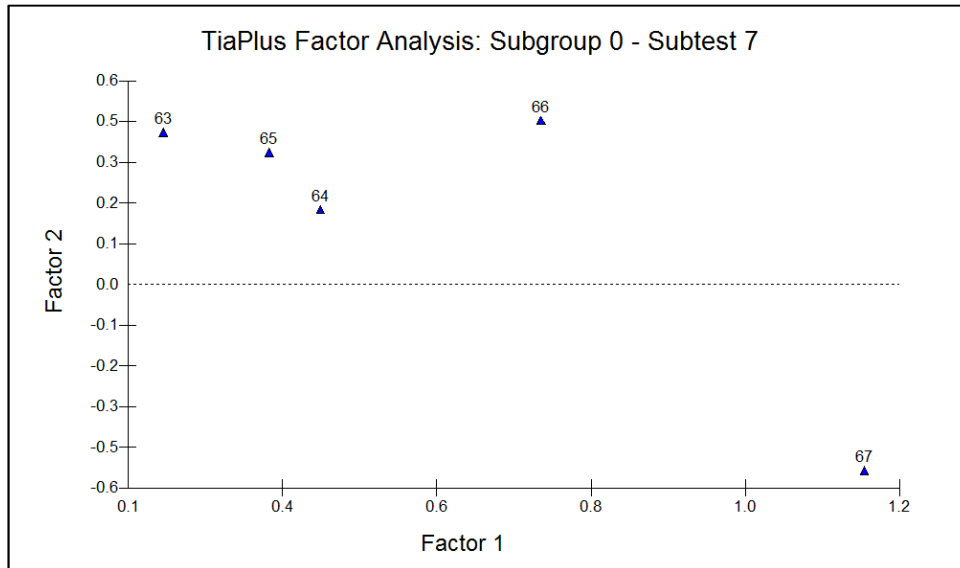
Die faktoranalises wat hieronder volg, wys ook hoedat die meerderheid items in elk van die betrokke subtoetse homogeniteit weerspieël, met die uitsondering van enkele uitskieters in elke subtoets. Soos vantevore vermeld is, behoort daar wanneer 'n ryk konstruk soos akademiese geletterdheid getoets word, voorsiening gemaak te word vir die teenwoordigheid van uitskiet-items in die ontwerp, mits hierdie items steeds 'n redelike mate van assosiasie met die res van die items vertoon. Uitskietitems soos Item 21 (kyk figuur 5.5) en Item 67 (kyk figuur 5.7), stip wel in elke geval weg van die groter groepering, maar korreleer steeds net met een faktor.



Figuur 5.5 Faktoranalyse: Woordeskat Enkelgaping (Tweede loodstoetsing)



Figuur 5.6 Faktoranalyse: Teksbegrip (Tweede loodstoetsing)



*Figuur 5.7 Faktoranalise: Vergelyking tussen tekste (Tweede loodstoetsing)*

Dit is ten slotte nodig om soos met die vorige toetsfase ook hierdie toetsfase se aantal potensiële misklassifikasies in oorweging te neem (kyk tabel 5.10). TiaPlus het die berekening van die misklassifikasiewaardes vir hierdie toetsfase egter slegs op die Alpha-koëffisiënt daarvan gebaseer, en nie ook op 'n GLB-betroubaarheidswaarde nie. Laasgenoemde kon nie vir hierdie toetsfase deur TiaPlus bepaal word nie, aangesien 'n groter studentekohort vir die berekening van die GLB-estimasië vereis word.

<b>Alpha-gebaseer:</b>	
Korrelasie tussen toets en hipotetiese, parallelle toets	8 (8.2%)
Korrelasie tussen waargenome en 'werklike' toetstellings	6 (5.9%)

*Tabel 5.10 Potensiële misklassifikasies van Toetsfase 2 soos bepaal deur TiaPlus*

Volgens TiaPlus se Alpha-gebaseerde berekening wissel die aantal potensiële misklassifikasies (vir die betrokke kohort van 99 studente) tussen 8 en 6, afhangende van watter korrelasiekriteria gebruik word. Omdat 50% van hierdie potensieel misgeklassifiseerde kandidate geslaag het terwyl hulle eintlik moes druij, beteken dit in hierdie geval dat 3 of 4 kandidate wat veronderstel was om te slaag, moontlik nie het nie. Hierdie statistiese bevinding hou in dat by die toekomstige administrasie van hierdie toetsfase die berekening van hoeveel

kandidate onder die geldende afsnypunt lê, op dieselfde manier gedoen kan word. Ter wille van regverdigheid, behoort sulke grensgevalle die geleentheid gegun te word om ook die derde en laaste skryftaak-assessering van hierdie meerfasige toets te voltooi.

#### **4. Slot**

In hierdie hoofstuk is daar verslag gedoen van die resultate van die loodstoetsing van die verfynde weergawe van TAGNaS. Die verfynde toets is saamgestel en geloods ooreenkomstig die spesifieke meerfasige ontwerp wat in Hoofstuk 3 vir hierdie assessering in die vooruitgesig gestel is. Op grond van die statistiese data wat aangebied is, asook die gepaardgaande interpretasies daarvan, is sekere gevolgtrekkings omtrent die produktiwiteit en betroubaarheid van die eerste en tweede veelkeusetoetsfases van TAGNaS gemaak. Samevattend, en oor die geheel geneem, blyk dit voorlopig uit hierdie hoofstuk dat beide toetsfases in hulle verfynde vorm as produktiewe en betroubare assesserings bestempel kan word wat beslis die potensiaal besit om resultate op te lewer waarvan die geldigheid van die interpretasies nie betwyfel hoef te word nie.

Daar is in terme van die primêr-geldende psigometriese kriteria vir hierdie studie (naamlik fasiliteitswaarde, diskriminasiewaarde, betroubaarheid en dimensionaliteit) aangetoon hoedat beide toetsfases en hulle onderskeie subtoetse in die meeste gevalle die vereiste statistiese standaard bereik. Waar skynbaar minder bevredigende resultate behaal is, is telkens kommentaar ter verklaring of ter regverdiging daarvan verskaf. Wat individuele items betref, word toegegee, en trouens aanbeveel, dat, soos wat die databank van itemstatistiek vir TAGNaS groei, daar in die toekoms verdere aanpassings en verfynings aan minder produktiewe items of probleemitems gemaak kan word.

Hoewel die grootte van die onderhawige studentekohort vir die doeleindes van hierdie hoofstuk bevredigend is, word verdere verfyning van individuele items



tydelik agterweë gehou totdat meer beslissende itemstatistiek van 'n nóg groter studentekohort verkry kan word. Op grond van die huidige datastel, kon potensiële individuele probleemitems wel geïdentifiseer word. Raadpleeg Bylaag E vir 'n lys van hierdie items en hulle prestasie.

'n Verdere belangrike gevolgtrekking wat op grond van die inligting in hierdie hoofstuk gemaak kan word, het te doen met die doelwit van hierdie studie om met gewysigde/nuwe subtoetse/taaktipes te eksperimenteer wat daartoe bydra dat die toetsontwerp beter met Patterson en Weideman (2013b) se vernude definisie van akademiese geletterdheid belyn. Sover aan my bekend, is hierdie die eerste ontwerppoging wat die gewysigde konstruk in die saamstel en implementering van 'n nuwe toets op die proef stel. Uit die data wat in hierdie hoofstuk, en die vorige een, aangebied is, wil dit lyk asof die eksperimentering met gewysigde subtoetse (soos die dubbele gaping-woordeskatsubtoets) en nuwe taaktipes (soos teksvergelyking, visuele opsomming en verwysings) wel geregverdig en die moeite werd was. Die statistiese data wat in hierdie studie gegeneer en gerapporteer is, dien as bewys dat hierdie subtoetse waardevolle nuwe toevoegings tot 'n toets van akademiese geletterdheid, soos TAGNaS, kan wees. Indien die huidige nagraadse toets in Engels (TALPS) ook op soortgelyke wyse verfyn en opnuut belyn kan word met die gemodifiseerde konstruk, behoort dit soortgelyke positiewe resultate op te lewer.

In die volgende en laaste hoofstuk word daar verder uitgebrei oor die gevolgtrekkings waartoe daar in hierdie studie gekom kan word, in die lig van die navorsingsdoelwitte wat in Hoofstuk 1 gestel is.

## Hoofstuk 6

### Gevolgtrekking

#### 1. Doel van die studie: 'n Samevattende terugblik

Ten spyte daarvan dat Afrikaans as akademiese taal toenemend in die huidige bestel onder druk verkeer, betoog hierdie studie aan die hand van verskeie navorsingsinsigte in Hoofstuk 1 dat daar wel plek is vir 'n Afrikaanse nagraadse toets van akademiese geletterdheid (TAGNaS) in die tersiêre taaltoetsomgewing, spesifiek as 'n alternatief vir die Test of Academic Literacy for Postgraduate Students (TALPS) wat slegs in Engels beskikbaar is. Du Plessis (2012) se bevindinge na aanleiding van die ontwerp van 'n tweede weergawe van TALPS was op hierdie punt onomwonde, en 'n mens moet aanvaar dat daar in sekere tradisioneel professionele en landbougerigte vakrigtings, soos onderwys, die regte, en weidingsleer en veeteelt, steeds 'n behoefte aan Afrikaans ook op nagraadse vlak sal bly. Gevolglik het hierdie ondersoek dit oorkoepelend ten doel gestel om in die behoefte aan só 'n toets te voorsien – 'n doelwit waaraan, in die lig van die strekking van die bewysvoering in die voorafgaande hoofstukke, hierdie studie wel suksesvol blyk te beantwoord het. Die bereiking van hierdie groter doelwit het egter ten nouste saamgehang met die verwesenliking van vier verdere, meer spesifieke doelwitte.

Vanselfsprekend oorskry 'n studie wat in die veld van die assessering van taalvermoë, en in hierdie geval 'n spesifieke soort taalvermoë, naamlik die beheersing van akademiese diskoers (Patterson & Weideman 2013a), gedoen word, die grense van nasionaal betekenisvolle tale, soos Afrikaans en Engels. Die studie se ander belangrike bydrae lê dus op daardie terrein, as 'n moontlike verruiming van ons insig oor hoe taaltoetse as toegepaste taalkundige meetinstrumente ontwerp en verfyn kan word. Sulke insigte is nie taalgebonden, maar bied nuwe ontwerpinsigte, of dui ten minste aan hoe vroeëre insigte meer geredelik verfyn kan word.

Daar word vervolgens samevattend aan die hand van kritiese nabetrugting aangetoon op watter manier, en in welke mate, hierdie studie daarin geslaag het om uitvoering te gee aan die spesifieke navorsingsdoelwitte wat aanvanklik gestel is.

### 1.1 Die ontwerp, loodsing en verfyning van TAGNaS

In die eerste plek het hierdie studie dit ten doel gehad om die ontwikkelingsgang van TAGNaS te dokumenteer. Daar is breedvoering in Hoofstuk 3, 4 en 5 aangetoon hoe die tipiese stappe in die toetsontwikkelingsproses (vgl. Bachman en Palmer, 1996:86-93, raadpleeg ook Read, 2010:290-293) vir hierdie toets verloop het. Soos gebruiklik in taaltoetsing, het die ontwerpproses begin met die artikulasie van die toetskonstruk (maar sien ook in die verband bespreking van die beperkings van die studie, hieronder), waarna dit geoperasionaliseer is deur middel van die ontwikkeling van toetsspesifikasies en die gepaardgaande taaktipes en items (Weideman, in voorbereiding).

Hierdie artikulasie van die konstruk en die daaropvolgende operasionalisering daarvan is, na die samestelling van die toets, gevolg deur twee rondtes van loodstoetsing op grond waarvan die geanaliseerde resultate as 'n wegwysers vir die evaluering en verfyning van die toets gebruik is. Gewoonlik dien die taaltoetsontwikkelingsproses egter slegs as agtergrond vir navorsing oor taaltoetse en word die suksesvolle uitvoering daarvan, opsigself beskou, nie altyd as 'n volwaardige navorsingsaktiwiteit in eie reg bestempel nie (Read, 2010:292), alhoewel hierdie siening wel betwis word (Weideman, Patterson & Pot, 2016; Rambiritch & Weideman, 2016). Die bereiking van hierdie navorsingsdoelwit is daarom nou verweef met, en wesenlik grondliggend aan, die bereiking van die volgende doelwit wat vir hierdie studie gestel is, naamlik die teoretiese begroning vir die ontwerp van TAGNaS.

## 1.2 Die teoretiese begroning vir die ontwerp van TAGNaS

Die vernaamste doelwit van hierdie studie was om aan te toon hoedat 'n teoreties-gefundeerde, wetenskaplik-verantwoordbare benadering tot toetsontwerp die ontwikkeling van TAGNaS ten grondslag lê. Die teoretiese begroning vir die ontwerp van hierdie toets is gebaseer op 'n reeks ontwerpbeginsels wat voorspruit uit Weideman (2017) se raamwerk van konstitutiewe en regulatiewe voorwaardes vir toegepaste linguistiese ontwerp (volledig behandel in Hoofstuk 2 van hierdie studie). Die ontwerpbeginsels het regdeur die toetsontwikkelingsproses as ontwerpmaatstaf gegeld, en waar van toepassing in die studie is gegewens oor die ontwerp en verfyning van TAGNaS dikwels aangebied met spesifieke verwysing na een of meer van die tegnies gestempelde ontwerpmomente in die raamwerk.

Ten einde 'n meer samehangende begrip te verkry van die verskillende opsigte waarin hierdie raamwerk die ontwerp van TAGNaS teoreties onderlê, is dit nodig om samevattend aan te dui hoedat, en in watter mate, die betrokke ontwerpbeginsels<sup>17</sup> in die uitvoering van hierdie studie gehandhaaf is. In die gedeelte wat volg, word die ontwerpbeginsel telkens eers in skuinsdruk verskaf, waarna kortliks aangedui word op watter manier dit onderliggend aan die ontwerp van TAGNaS is.

*Integreer, op 'n sistematiese manier, veelvuldige bewyse in die formulering van 'n geldigheidsargument vir 'n toets.* Vir toetsontwikkelaars om hulle interpretasies van toetsresultate te regverdig, moet 'n geldigheidsargument opgebou word. Volgens Van der Walt en Steyn (2007:142) neem 'n geldigheidsargument die vorm aan van 'n reeks aansprake of hipoteses (met geïmpliseerde teenaansprake) tesame met relevante bewyse as ondersteuning daarvoor. Hierdie bewyse kan bestaan uit beide teoretiese redenasies en verskeie soorte empiriese bewyse verkry van loodstoetsing (Read, 2010:288). Hoewel

---

17. Vir 'n skematiese opgaaf van die ontwerpbeginsels raadpleeg weer tabel 2.8 in Hoofstuk 2.

hierdie studie in sy huidige formaat nie formeel die vorm aanneem van 'n koherente geldigheidsargument vir TAGNaS nie, verskaf dit nietemin reeds 'n geïntegreerde opgaaf van veelvuldige teoretiese en empiriese bewyse wat in die formulering van sodanige argument ingespan sou kon word. Daar is byvoorbeeld aangetoon dat

- 'n Afrikaanse eweknie vir TALPS oor 'n hoër vlak van gesigsgeldigheid onder Afrikaanse studente sal beskik;
- die toets gebaseer is op 'n teoreties verdedigbare konstruk;
- die verfynde toetsfases oor hoogs bevredigende algehele betroubaarheidskoëffisiënte beskik, selfs wanneer van die mees konserwatiewe indekse soos Cronbach Alpha gebruik gemaak word (Toetsfase 1 = 0.86; Toetsfase 2 = 0.92);
- die items wat in die verfynde toets ingesluit is se fasiliteits- en diskriminasiewaardes, asook “Infit MNSQ”-waardes (waar van toepassing) binne die voorkeurparameters val; en
- die interne korrelasies tussen die subtoets van die verfynde toets 'n aanvaarbare vlak van dimensionele samehang vertoon volgens sekere kriteria.

Soos wat verdere loodsing van TAGNaS, en uiteindelik ook die administrasie daarvan, plaasvind, kan hierdie lys van bewyse aangevul en uitgebrei word vir die doeleindes van 'n formele geldigheidsargument.

*Stip duidelik aan en verduidelik aan die publiek die toepaslik beperkte reikwydte van die toets, asook die grense en beperkings van sy meting, en doen dit op 'n beskeie manier.* In hierdie studie is daar van die staanspoor af aangedui dat TAGNaS as 'n potensiële Afrikaanse alternatief vir TALPS bedoel word. Die toets het dus beperkte metingsomvang; dit is gebonde aan die konstruk wat gemeet word, wat spesifiek bedoel is om akademiese geletterdheidsvlakke te bepaal, en die betrokke teikengebruiker is voornemende Afrikaanse nagraadse studente.

*Verseker dat die toetsresultate wat verkry word, toereikend konsekwent is, ook met die verloop van tyd, en oor verskeie administrasies heen.* Hoofstuk 4 en 5 van hierdie studie sluit verskeie betroubaarheidsanalises in wat na afloop van die twee toetsloodings op die onderskeie subtoetse en die toetsfases in die geheel uitgevoer is. Die Alpha- en GLB-waardes wat in hierdie studie verkry is, was oorwegend bevredigend. Waar skynbaar minder bevredigende waardes behaal is, is telkens kommentaar ter verklaring of ter regverdiging daarvan verskaf. Die twee verfynde toetsfases beskik oor hoogs bevredigende algehele betroubaarheidskoeffisiënte, onderskeidelik 0.86 en 0.92, wat heelwat hoër as die kerf is wat internasionaal vir hoë impak toetse gestel word.

*Verseker doeltreffende meting deur van 'n meetinstrument gebruik te maak waarvan die toereikendheid (geldigheid) verdedig kan word.* Die lys van teoretiese en empiriese bewyse wat by die eerste ontwerpbeginsel hierbo verskaf is, geld ook vir hierdie een. Dit dien as verdediging vir die geldigheid van die interpretasies van die toetsresultate.

'n Toets is in die eerste plek 'n geldige instrument omdat dit aangewend kan word om 'n meting te verskaf. Die meting dien as oorsaak van die meetresultaat. Daardie resultaat is op sy beurt die gevolg van die aanwending van die toets. Hierdie relasie tussen oorsaak en gevolg is 'n eerste aanduiding van die geldigheid van die toets; dit wys dat die toets se werking wel geskied het. Vanselfsprekend is hierdie nog nie voldoende of toereikende aanduidings van geldigheid nie. Eers wanneer die verruiming van hierdie vlak van geldigheid gedoen word deur verdere vrae te vra, kry ons meer bevredigende aanduidings van toetsgeldigheid. Hierdie verdere vrae word hier onder bespreek, en behels o.a. die vraag: “Hoe definieer ons wat gemeet word?” (m.a.w. die konstruk) en die verdere vraag: “Hoe dra ons insig tot wat gemeet word, ons begrip van die konstruk, by tot die betekenisvolle interpretasie van die resultate?” Beide hierdie verdere vrae behels verruimde momente van die oorspronklike kousale vraag oor

toetsgeldigheid (kousaal omdat dit afleibaar is uit die tegniese werking van die toets as oorsaak van gevolge, die resultate).

*Gebruik 'n toets wat toepaslik en toereikend gedifferensieerd is, veral wat betref die tipe en verskeidenheid subtoetse.* Soos duidelik uit Hoofstuk 3 na vore gekom het, is akademiese geletterdheid 'n veelvlakkige konstruk waarvan die multidimensionaliteit nie vasgevang sou kon word in 'n beperkte aanbod van gelyksoortige subtoetse en/of taaktipes nie. In die ontwerp van TAGNaS is daar daarom voorsiening gemaak vir die insluiting van tien verskillende subtoetse in 'n meerfasige toetsformaat waarvan sommige boonop uit nuut of gewysigde taaktipes bestaan. Die gedifferensieerdheid van die toetsontwerp doen nietemin nie afbreek aan die homogeniteit van die toets in terme van dit wat gemeet word nie, soos wat die interne korrelasie- en faktoranalises in Hoofstuk 4 en 5 bewys het. Die stelling kan gewaag word dat die sukses van TAGNaS en soortgelyke toetse van akademiese geletterdheid, veral wanneer dit vergelyk word met ander toetse wat op dieselfde konstruk gebou is, maar meer monotoon in hul ontwerp is (Myburgh, 2015; Sebolai, 2016), grootliks toegeskryf kan word uit die meer verbeeldingryke verskeidenheid van subtoetse wat hul ontwerp uitmaak. Die beginsel van diversifikasie van toetstake en selfs -fases word klaarblyklik deur hierdie studie se resultate bevestig.

*Maak die toets intuïtief aanneemlik en aanvaarbaar vir die gebruikers daarvan (oorweeg die gesigsgeldigheid van 'n toets).* Wat die gesigsgeldigheid van TAGNaS betref, het spesifiek twee oorwegings in hierdie studie op die voorgrond getree. In die eerste plek is daar, as deel van die motivering vir hierdie studie, aangevoer dat 'n Afrikaanse nagraadse toets van akademiese geletterdheid waarskynlik gunstiger deur Afrikaanssprekende studente ontvang sal word as 'n Engelse teenvoeter – 'n afleiding wat gemaak is op grond van die bevindinge van Du Plessis (2012) in haar resepsiestudie van TALPS in 2012. Tweedens, is daar, met inagneming van Butler (2009) se bevindinge oor persepsies omtrent nagraadse studente se akademiese skryfvermoë, en hoe dit

getoets behoort te word, besluit om veral ter wille van die beginsel van aanneemlikheid 'n akademiese skryfoefening by die derde fase van TAGNaS se ontwerp in te sluit. Soortgelyke inisiatiewe ter verbetering van die gesigsgeldigheid van hierdie tipe toetse is ook te vinde by die ontwerp deur Drennan (2017) en 'n span toetsontwerpers van 'n toets van skryfgereedheid, die Assessment of Preparedness to Produce Multimodal Information (APPMI), wat beoog om die pre-skryfgereedheid van senior studente in die sosiale wetenskappe te meet. Die ontwerpers van APPMI het reeds van die insigte van hierdie studie gebruik gemaak, en baat gevind by die eksperimentering met nuwe taaktipes, wat met APPMI verder gevoer gaan word.

*Verskaf in die mees resente terme 'n teoretiese verdediging van dit wat getoets word.* In Hoofstuk 3 van hierdie studie is indiepte besonderhede verskaf omtrent die teoretiese rasionaal wat as basis vir die ontwikkeling van TAGNaS gebruik is. Die ontwerp van hierdie toets berus op 'n onlangs hersiene konstruk van akademiese geletterdheid (Patterson & Weideman, 2013b) wat veral die analitiese karakter van akademiese diskoers beklemtoon. Juis omdat die konstruk van die toets noukeurig geartikuleer is, bied dit, op 'n volgende vlak van geldigheid, toereikende gronde vir die regverdiging van die inhoud van die vermoë wat gemeet word: die vermoë om akademiese diskoers te hanteer. Verder, juis omdat die konstruk hierdie tipies analities-gekwalfiseerde taalvermoë voldoende artikuleer, en met verwysing na meer resente, funksioneel gedefinieerde perspektiewe op wat taalvermoë behels, is die verdere operasionalisering daarvan ook teoreties verdedigbaar. Die verruiming van die aantal en tipe subtoetse moet uiteindelik krities beoordeel word in die lig van die teoretiese rasionaal van die toets, die konstruk. Die bespreking van hierdie sake in die studie bied goeie gronde vir die stelling dat die teoretiese regverdiging, en die operasionalisering van die konstruk vir TAGNaS, voldoen aan hierdie ontwerpvoorwaarde.



*Maak seker dat die toets interpreteerbare en betekenisvolle resultate oplewer.*

Om die nakoming van hierdie vereiste te bewerkstellig, is verskeie belangrike stappe, waarna reeds hierbo verwys is, in die ontwerp van TAGNaS gevolg. Dit sluit die volgende in:

- die toets is gebaseer op 'n teoreties verdedigbare konstruk;
- toetsspesifikasies is saamgestel wat die konstruk met die toetsitems bely;
- die keuse van die leestekste is uitgevoer ooreenkomstig bepaalde kriteria;
- verskeie lede van die ontwerpspan wat gebruik is, beskik oor bewese praktiese ontwerpervaring van akademiese geletterdheidstoetse;
- twee loodstoetsrondtes is (tot dusver) uitgevoer vir die doeleindes van verfyning;
- die loodstoetsresultate is statisties geanaliseer ten einde die produktiwiteit daarvan te bepaal; en
- verskeie bewyse (afgelei van die statistiese analises) is verskaf wat die geldigheid van die interpretasies van die loodstoetsresultate bevredigend ondersteun.

Die uiteindelijke voldoening van TAGNaS aan hierdie ontwerpvoorwaarde sal egter kom eers wanneer dit ten volle geïmplementeer en aangewend word om die risikovlakke van voornemende nagraadse studente ten opsigte van taal te bepaal. In daardie geval, soos in die geval van TALPS en ander, soortgelyke toetse, sal empiriese gegewens verder aangewend word om verskillende risikovlakke te identifiseer en te benoem (bv. van Geen risiko of Min tot geen risiko tot by Uiters hoë risiko).

*Maak die toets, asook inligting omtrent die toets, toeganklik vir almal.* Wanneer die administrasie van TAGNaS 'n aanvang sou neem, kan 'n platform soos die ICELDA-webwerf gebruik word om die toets, asook inligting omtrent die toets, maklik bereikbaar vir gebruikers te maak. Sien ook die bespreking hier onder by “Wees bereid om aan die publiek verslag te doen oor hoe die toets gebruik is of gebruik gaan word”.

*Verkry die toetsresultate op 'n effektiewe manier en verseker dat dit bruikbaar is.* In 'n poging om TAGNaS meer effektief te operasionaliseer en administreer, is die grondslag in hierdie studie gelê vir die implementering van 'n potensiële driefasetoetsontwerp, 'n formaat wat byvoorbeeld die gebruik van 'n verpligte/opsionele siftingstoets 'n haalbare assesseringsmoontlikheid maak. Die empiriese bewyse wat binne die bestek van hierdie studie gegenerer kon word, sal egter verder aangevul moet word vir meer besliste gevolgtrekkings omtrent die effektiwiteit en bruikbaarheid van 'n siftingsmeganisme of eerste toetsfase.

*Belyn die toets met die onderrig wat óf daarop volg óf dit voorafgaan, en so na as moontlik met die leerproses.* Die belyning van TAGNaS spesifiek met die onderrig-en-leerproses val buite die omvang van hierdie studie. Waar die belyningsbeginsel egter wel in die ontwerp van TAGNaS ter sprake gekom het, was met die samestelling van die toetsspesifikasies. Soos wat tradisioneel in taaltoetsontwikkeling vereis word, is 'n bloudruk vir TAGNaS geskep met die doel om belyning tussen die verskillende komponente van die toets (toetskonstruk, subtoetse, toetsitems) te verseker. Raadpleeg bylaag B vir 'n volledige uiteensetting van die toetsspesifikasies.

*Wees bereid om aan die publiek verslag te doen oor hoe die toets gebruik is of gebruik gaan word.* Die uitvoering van hierdie studie kan beskou word as die eerste stap in die proses om deursigtig te wees, en gevolglik sosiaal verantwoordelik teenoor die voornemende gebruikers van die toets op te tree. Wat toeganklikheid betref, bied die insigte van Rambiritch (2010) en Rambiritch en Weideman (2016) 'n eerste aanduiding van hoe informasie oor die aard en inhoud van 'n taaltoets verantwoordelik openbaar gemaak kan word.

*Heg waarde aan die integriteit en kwaliteit van die toets. Maak geen toegewings wat die toets se status as regverdige meetinstrument sal ondermyn nie.* Die versameling teoretiese en empiriese bewyse wat in hierdie studie aangebied is, dien tot staving van die handhawing van hierdie ontwerpbeginsel. Waar

toegewings tussen botsende ontwerpvereistes wel in die ontwikkelingsproses gemaak moes word (Weideman, 2017), is daar telkens genoegsame motivering verskaf om die redelikheid daarvan te verdedig.

Die voldoening aan hierdie vereiste vir taaltoetsontwerp kan eintlik nie genoeg beklemtoon word nie, soos Du Plessis (2017) se doktorsale studie ook so pas aangetoon het. In haar studie is die uiters onbevredigende manier waarop tuistale in die finale eksamen aan die einde van graad 12 geëksamineer word, aan die orde gestel, en bied haar bevindings 'n verklaring vir die wye publieke wantroue in die resultate van hierdie eksamens se uitslae. As die integriteit van die eksamen aangetas word deur meting van swak of minderwaardige kwaliteit te verdra, ondermyn dit sonder twyfel enige toets se status as regverdige meetinstrument.

Dit is verder duidelik uit die voorafgaande dat tegniese sorgsaamheid voorop gestel is in die ontwerp van die toets. Daar is verskeie kere in die voorafgaande bespreking aangetoon hoe bv. potensiële misklassifikasies menslik hanteer kan word, deur aan studente wat die toets aflê en grensgevalle uitmaak, of 'n tweede kans of 'n vrystelling van een van die fases van die toets te bied, ten einde almal so regverdig moontlik te behandel. Die wyse waarop die toets se metings in onderskeie fases gedoen sal kan word, dra natuurlik verder by tot die meer algemene bevrediging van hierdie ontwerpvoorwaarde van tegniese regverdigheid en billikheid.

*Onsien geen moeite om die toets toepaslik vertrouens- en agtenswaardig te maak nie.* Die noukeurigheid waarmee daar gepoog is om aan die bostaande raamwerk van ontwerpbeginsels getrou te bly, is bewys daarvan dat verantwoordelike toetsontwerp hoog op prys gestel is in die uitvoering van hierdie studie. Sulke sorgsaamheid bied 'n basis waarop 'n toets uiteindelik 'n reputasie kan verkry van 'n geloofbare meetinstrument, een waarvan die resultate jy op kan staatmaak. Die groot prys vir enige toets is dat dit geag word en

tegniese vertrouwe kan ontlok, en so ook aan hierdie ontwerpvoorwaarde kan voldoen. In hierdie opsig staan TAGNaS nog aan die begin.

### 1.3 Die eksperimentering met nuwe en/of gewysigde subtoetse/taaktipes

Aangesien die ontwerp van TAGNaS, anders as dié van TALPS, op Patterson en Weideman (2013b) se bygewerkte konstruk van akademiese geletterdheid gebaseer is, was gepaardgaande aanpassings ten opsigte van die operasionalisering daarvan 'n noodwendige gevolg. Die derde doelwit van hierdie studie was daarom om, in ooreenstemming met die hersiene toetskonstruk, sekere gewysigde en/of nuwe subtoetse/taaktipes wat veral logiese en/of analitiese denke betrek, vir die eerste keer in TAGNaS op die proef te stel.

Gebaseer op die statistiese data in Hoofstuk 5 is daar tot die gevolgtrekking gekom dat hierdie eksperimentering met die aangepaste en nuwe subtoetse nie slegs geregverdig en die moeite werd was nie, maar as suksesvol en in sekere opsigte baanbrekend was. Tegniese vernuwing is nie altyd dramaties en abrupt nie, maar dikwels stadig en inkrementeel (Weideman, 2017). Daar is in hierdie studie aangetoon dat subtoetse soos die dubbele gaping-woordeskatsubtoets, teksvergelyking, visuele opsomming en verwysings waardevolle nuwe toevoegings tot 'n toets van akademiese geletterdheid kan wees.

### 1.4 Die moontlike implementering van 'n driefasige toetsformaat

As 'n vierde en laaste doelwit het hierdie studie dit as oogmerk gehad om 'n voldoende aantal produktiewe toetsitems te skep ten einde eksperimentering met 'n driefasetoetsmodel moontlik te maak. Hierdie doelwit is wel bereik en 'n verfynde weergawe van TAGNaS waarvan die subtoetse en toetsitems in drie fases opgedeel is, is uiteindelik as eindproduk voortgebring. Die eerste twee toetsfases behels die veelkeusetoetsmetode, terwyl die derde en laaste fase uit 'n

akademiese skryftaak bestaan. Die eerste toetsfase word as 'n opsionele/verpligte siftingstoets beoog wat dit moontlik maak om sterker, meer bedrewe kandidate vroeër uit die toetsingsproses uit te skakel.

## **2. Beperkings van die studie**

Die loodsing van die toetsitems vir TAGNaS moes binne 'n bepaalde tydsbestek afgehandel word ten einde hierdie studie binne die voorgeskrewe periode te kon voltooi. Daar was dus beperkinge wat betref die aantal beskikbare geleenthede vir loodstoetsing, asook die grootte van die loodstoetspopulasies wat vir die studie byeengeroep kon word. Verder was dit prakties onmoontlik om 'n groot genoeg kohort van slegs nagraadse of voornemende nagraadse Afrikaanse studente vir die loodstoetsing byeen te kry wat boonop gewillig sou moes wees vir deelname aan meer as een toetssessie. Daar is daarom van 'n gemengde groep voorgraadse studente gebruik gemaak wat van eerste- tot derdejaarsvlak gewissel het.

Die data-analises in hierdie studie is uitgevoer primêr vanuit die klassieke toetsteorie-benadering (KTT). 'n Welbekende leemte in hierdie benadering is dat die resultate van die analises potensieel 'n mate van onbestendigheid kan vertoon wanneer verskillende toetspopulasies by dieselfde ondersoek betrek sou word. Die rede hiervoor is dat, soos vroeër reeds vermeld is, 'n toets se eienskappe in hierdie benadering nie van die eienskappe van die kandidate wat dit aflê, geskei kan word nie. Om vir hierdie tekortkoming te probeer vergoed, is daar in hierdie studie ook, waar moontlik, van Rasch-analise gebruik gemaak wat op item-responsteorie (IRT) gebaseer is. Die beperking op IRT-tipe analises, wat van waarskynlikheidsberekeninge gebruik maak, is egter dat dit slegs op grond van veel groter getalle toetslinge se resultate gedoen kan word as wat in alle gevalle vir TAGNaS se loodsing beskikbaar was. Nogtans is die analises in die enkele gevalle waar dit wel uitgevoer kon word, amper volledig in ooreenstemming met

die bevindinge wat met KTT-tipe analyses verkry is, veral ten opsigte van die kwaliteit van items.

'n Verdere moontlike tekortkoming het aan die lig gekom deur die loop van die studie. Dit hou verband met sake wat aan die begin van hierdie hoofstuk aan die orde gestel is met betrekking tot die posisie van Afrikaans op tersiêre en veral nagraadse vlak, en het te doen met die konseptualisering van die proses van toetsontwerp. Waar Fulcher (2010:94) die proses se begin beskou as die artikulering van die toetskonstruk, 'n siening wat breedweg ook deur hierdie studie onderskryf is, blyk dit dat daar wel meer genuanseerde perspektiewe op hierdie saak moontlik is. So bv. wys Weideman (in voorbereiding) daarop dat daar in werklikheid dikwels twee of drie fases in die ontwerpproses is wat konstruk-besluite voorafgaan. Hoekom dit wel ter sake is ook in die ontwerp van TAGNaS dat die besluit om in meer as een akademiese taal te toets, of slegs in een, sonder uitsondering een is wat die politieke aard van sulke besluite bevestig.

Die amptelike sanksionering van toetsontwerpers se werk kan dus heelwat vroeër in die proses al die toon aangee en die proses in 'n rigting stuur wat nie altyd verantwoordelike taaltoetsontwerp bevorder nie. Ter wille van die fokus van hierdie studie is hierdie moontlike beperking nie verder ondersoek nie. Sou die implementering en gebruik van TAGNaS egter wel gerade blyk te wees, sal al hierdie oorwegings beslis aan die orde moet kom. Ten einde verantwoordelik te toets, moet daar verantwoordelike besluite geneem word binne die institusionele verband waarin die toetsing gedoen moet word. Sonder behoorlike institusionele besluitneming is enige taalintervensie, of dit nou beleidmatig van aard is, of 'n taalkursus of taaltoets is, eintlik van meet af aan gedoem.

Die finale moontlike beperking van die studie wat hier bespreek word, het te doen met die gebruikmaking van 'n teoretiese raamwerk wat spruit uit die werk van Weideman (soos saamgevat in Weideman, 2017). Terwyl daar in hierdie studie ruimskoots verwys is na ander relevante diskussies en gesaghebbendes,

veral in die bespreking van die resente en vroeëre literatuur, is die werk wat afkomstig is uit hierdie teoretiese oord (soos ook beproef in die veld van taaltoetsing deur veral Van Dyk, 2010 en Rambiritch, 2012) uitgesonder as die grondliggende stel onderskeidings en die basis van die ondersoek. Die rede waarom daar so sterk aangeleun is op hierdie raamwerk behoort egter duidelik te wees. In 'n studie van beperkte omvang soos hierdie is dit gerade, en moontlik meer produktief, om eerder 'n enkele teoretiese uitgangspunt na behore uit te pluis en die robuustheid daarvan te beproef as om te poog om diverse standpunte te probeer sintetiseer. In elk geval bied die teoretiese raamwerk vir hierdie studie op sigself so 'n platform waarop ander standpunte geweeg en beoordeel kan word, en na waarde geskat en geïnterpreteer kan word. Laastens is dit, sover my kennis strek, die enigste huidige poging om 'n volwaardige teorie van die toegepaste taalkunde te ontwikkel, wat amper 'n noodsaaklikheid is in 'n veld wat erg teenstrydige en konflikterende benaderingswyses huisves. In hierdie studie het die produktiewe aanwending van hierdie uitgangspunt dus voorrang geniet, en hopelik ook sonder om ander alternatiewe uit te skakel of te ignoreer.

### **3. Aanbevelings vir verdere navorsing**

Benewens die noodsaaklikheid van verdere loodstoetsing, verfyning en validering, ontsluit hierdie studie verskeie ander verwante navorsingsmoontlike wat in die toekoms ontgin sou kon word. In die eerste plek kan die effektiwiteit van die meerfasige toetsmodel in daaropvolgende studies ondersoek word. Geldenhuys (2007) stel in hierdie verband die uitvoering van verskeie korrelasiestudies voor. Soos wat verdere data beskikbaar kom, kan prestasie in die siftingstoets met prestasie in die daaropvolgende tweede en/of derde toetsfases van TAGNaS gekorreleer word ten einde die voorspellingsgeldigheid daarvan beter te bepaal of te bevestig. Indien dit suksesvol uitgevoer kan word, kan die driefase-model wat vir TAGNaS voorgestel is, natuurlik na ander toetse, soos TALPS en APPMI, uitgebrei word.

Verder is die belyning van 'n konstruk met geskikte taaktipes nooit 'n afgehandelde onderneming nie. Toetskwaliteit is, in die woorde van Weideman (2014), in groot mate afkomstig van verbeeldingryke omgang met die toetskonstruk en die ontwerp van 'n toets op basis daarvan. Verdere studies kan dus daarop gemik word om die konstruk van akademiese geletterdheid waarop TAGNaS gebaseer is, nog meer sorgsaam en berekend te omskryf, en om toenemend vernuwend en kreatief te werk te gaan met die ontwerp van taaktipes wat met die komponente daarvan belyn. TAGNaS het alreeds gedien, soos hierbo vermeld is, om die eksperimentering met nuwe subtoetse in 'n nuwe toets, APPMI (Drennan, 2017) te stimuleer.

Alhoewel TAGNaS net soos TALPS nie as 'n vakspesifieke akademiese geletterdheidstoets bedoel is nie, is daar in die nagraadse universitêre omgewing, veel meer as op voorgraadse vlak, 'n sterk saak uit te maak vir vak- of dissipline-spesifieke taaltoetsing. APPMI wil bv. dien om die akademiese geletterdheid van senior voorgraadse en junior nagraadse studente in die sosiale wetenskappe verantwoordelik te meet. Toetsontwikkelaars in akademiese geletterdheid kan hulle daarom toenemend op die ontwerp en teoretiese begroning van sodanige toetse toespits, en later van tyd selfs studies onderneem wat daarop gemik is om die mate van ekwivalensie tussen hierdie toetse se metings te bepaal.

Laastens, omdat daar met hierdie studie gepoog is om 'n bydrae te lewer tot die bevordering van veeltaligheid op tersiêre vlak word die verwagting hoopvol gekoester dat dit die ontwerp van nagraadse akademiese geletterdheidstoetse in ook van die ander amptelike landstale sal stimuleer. Die eksperimentering hiermee is tans nog kleinskalig, en nie prominent genoeg gerapporteer nie, maar dit verdien wel verdere en meerdere aandag.

Ten spyte van goeie werk in die verlede, is die veld van taalassessering in Suid-Afrika eintlik nog in sy kinderskoene. Daar is wel tekens (sien bv. die inisiatiewe aangeteken op ICELDA 2015) dat dit groter prominensie begin verkry, wat 'n



hoopvolle teken is. Hierdie studie poog om 'n beskeie maar tog sinvolle bydrae tot hierdie ontwikkeling te lewer.

## **Bronnelys**

- Alderson, J.C., Clapham, C. & Wall, D. 1995. *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- American Educational Research Association. 1999. *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- Bachman, L.F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F. 2004. *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Bachman, L.F. & Palmer, A.S. 1996. *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L.F. & Palmer, A.S. 2010. *Language assessment in practice*. Oxford: Oxford University Press.
- Bachman, L.F. & Purpura, J.E. 2010. Language assessments: Gate keepers or door openers? In Spolsky, B. & Hult, F.M. *The handbook of educational linguistics*. Chichester, West Sussex: Wiley-Blackwell. 456-468.
- Blanton, L.L. 1994. Discourse, artefacts and the Ozarks: Understanding academic literacy. *Journal of Second Language Writing*, 3(1):1-16.
- Bornman, E., Pauw, J.C. & Potgieter, P.H. 2014. Houdings en opinies oor moedertaalonderrig en die keuse van 'n universiteit: Afrikaanssprekende studente aan Unisa. *Tydskrif vir Geesteswetenskappe*, 54(4):596-609.

- Brown, J.D. 2004. Research methods for applied linguistics: Scope, characteristics and standards. In Davies, A. & Elder, C. (reds.). *The handbook of applied linguistics*. Malden, MA: Blackwell. 476-500.
- Brumfit, C. 1995. Teacher professionalism and research. In Cook, G. & Seidlhofer, B. (reds.). *Principle and practice in applied linguistics*. Oxford: Oxford University Press. 27-42.
- Butler, G. 2007. A framework for course design in academic writing for tertiary education. Ongepubliseerde doktorsproefskrif, Universiteit van Pretoria. Beskikbaar by <http://hdl.handle.net/2263/27887>.
- Butler, G. 2009. The design of a postgraduate test of academic literacy: Accommodating student and supervisor perceptions. *Southern African Linguistics and Applied Language Studies*, 27(3):291-300.
- Butler, G. 2011. Supervisor perceptions of the academic literacy requirements of postgraduate students at the University of Pretoria. *Tydskrif vir Taalonderrig*, 45(1):7-22.
- Carroll, J.B., Carton, A.S. & Wilds, C.P. 1959. *An investigation of cloze items in the measurement of achievement in foreign languages*. Cambridge, MA: College Entrance Examination Board.
- Chapelle, C.A. 2012. Conceptions of validity. In Fulcher, G. & Davidson, F. (reds.). *The Routledge handbook of language testing*. Londen: Routledge. 21-33.
- Cito. 2005. *TiaPlus user's manual*. Arnhem: Cito Measurement and Research Department.
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly*, 34(2): 213-238.

- Cronbach, L.J. & Meehl, P.E. 1955. Construct validity in psychological tests. *Psychological Bulletin*, 52:281-302.
- Crystal, D. 1987. *The Cambridge encyclopaedia of language*. New York: Cambridge University Press.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. & McNamara, T. 1999. *Dictionary of language testing*. (Reeks-redakteur: Milanovic, M.). Cambridge: Cambridge University Press.
- Davies, A. & Elder, C. 2004. General introduction. Applied linguistics: Subject to discipline? In Davies, A. & Elder, C. (reds.). *The handbook of applied linguistics*. Malden, MA: Blackwell. 1-15.
- Davies, A. & Elder, C. 2005. Validity and validation in language testing. In Hinkel, E. (red.). *Handbook of research on second language teaching and learning*. Mahwah, New Jersey: Lawrence Erlbaum Associates. 795-813.
- DELNA. 2016. Diagnostic English language needs assessment. [Aanlyn]. Beskikbaar by <http://www.delna.auckland.ac.nz/en.html>. [20 Augustus 2016].
- Die Afrikaanse Taalraad. 2016. *Taalraad-persverklaring: Taalbeleid by die US, UP en UNISA*. 23 Junie. [Aanlyn]. Beskikbaar by <http://www.litnet.co.za/atr-persverklaring-taalbeleid-die-us-en-unisa/>. [8 Desember 2016].
- Douglas, D. 2010. *Understanding language testing*. Londen: Hodder Education.

- Drennan, L. 2017. Defensibility and accountability: Developing a theoretically justifiable academic writing intervention for Sociology students at tertiary level. Ongepubliseerde en onvoltooide doktorsproefskrif, Universiteit van die Vrystaat.
- Du Plessis, C.L. 2012. The design, refinement and reception of a test of academic literacy for postgraduate students. Ongepubliseerde meestersgraadverhandeling, Universiteit van die Vrystaat.
- 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 examinations. Ongepubliseerde doktorsproefskrif, Universiteit van die Vrystaat.
- Farhady, H. 1979. The disjunctive fallacy between discrete-point and integrative tests. *TESOL Quarterly*, 13(3):347-357.
- Fulcher, G. 1997. Text difficulty and accessibility: Reading formulae and expert judgement. *System*, 25(4):497-513.
- Fulcher, G. 2010. *Practical language testing*. Londen: Hodder Education.
- Fulcher, G. & Davidson, F. 2007. *Language testing and assessment. An advanced resource book*. Londen: Routledge.
- Gee, J.P. 1998. What is literacy? In Zamel, V. & Spack, R. (reds.). *Negotiating academic literacies: Teaching and learning across languages and cultures*. Mahwah, New Jersey: Lawrence Erlbaum Associates. 51-59.
- Geldenhuys, J. 2007. Test efficiency and utility: Longer or shorter tests. *Ensovoort*, 11(2):71-82.

- Gericke, M. 2016. Akademia in Bfn oorval met navrae. 14 Oktober. [Aanlyn].  
Beskikbaar by <http://www.netwerk24.com/Nuus/Onderwys/akademia-in-bfn-oorval-met-navrae-20161013>. [8 Desember 2016].
- Giezenaar, G. & Schouten, E. 2002. *Wijze woorden: Woordenlijst academisch Nederlands*. Amsterdam: Intertaal.
- Gous, N. 2016. Unisa nou ook nét Engels. 23 Junie. [Aanlyn]. Beskikbaar by  
<http://www.netwerk24.com/Nuus/Onderwys/unisa-nou-ook-net-engels-20160623>. [8 Desember 2016].
- Green, A. 2014. *Exploring language assessment and testing. Language in action*.  
Londen: Routledge.
- Green, R. 2013. *Statistical analyses for language testers*. New York: Palgrave  
Macmillan.
- Guyer, R. & Thompson, N.A. 2011. *User's manual for IteMan 4.2*. St Paul,  
Minnesota: Assessment Systems Corporation.
- Hall, C.J., Smith, P.H. & Wicaksono, R. 2011. *Mapping applied linguistics. A guide  
for students and practitioners*. New York: Routledge.
- Hay, HR. 2010. Korrespondensie: Kantoor van die viserektor. 18 Augustus,  
Universiteit van die Vrystaat.
- Hughes, A. 2003. *Testing for language teachers*. Cambridge: Cambridge  
University Press.

- Hymes, D.H. 1972. On communicative competence. In Pride, J.B. & Holmes, J. (reds.). *Sociolinguistics: Selected Readings*. Harmondsworth, VK: Penguin. 269-293.
- ICELDA (Inter-Institutional Centre for Language Development and Assessment). 2015. Language test development and research. [Aanlyn]. Beskikbaar by <http://icelda.sun.ac.za/>. [6 September 2015].
- Jansen, L. 2013. More academic filters in waiting for candidates. *Star*, 4 Januarie. 5.
- Johns, A.M. 1997. *Text, role, and context. Developing academic literacies*. Cambridge: Cambridge University Press.
- Le, P.L. 2011. Assessing academic literacy of first year Vietnamese university students: How appropriate is the TALL? Ongepubliseerde meestersgraadverhandeling, Universiteit van Groningen.
- Le, P.L., Du Plessis, C. & Weideman, A. 2011. Test and context: the use of the test of academic literacy levels (TALL) at a tertiary institution in Vietnam. *Tydskrif vir Taalonderrig*, 45(2):115-13.
- Lea, M.R. & Street, B.V. 1998. Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2):157-172.
- McNamara, T. & Roever, C. 2006. *Language testing: The social dimension*. Oxford: Blackwell.

- Maroela Media, 2016. Talle student wil gevorderde grade in Afrikaans doen. 12 September. [Aanlyn]. Beskikbaar by <http://maroelamedia.co.za/goeiegoed/goeie-nuus/talle-studente-wil-gevorderde-grade-in-afrikaans-doen/>. [8 Desember 2016].
- McNamara, T. 2000. *Language testing*. Oxford: Oxford University Press.
- McNamara, T. 2011. Language assessment as social practice. In Wei, L. (red.). *The Routledge applied linguistics reader*. Londen: Routledge. 353-363.
- Messick, S. 1980. Test validity and the ethics of assessment. *American Psychologist*, 35(11):1012-1027.
- Messick, S. 1981. Evidence and ethics in the evaluation of tests. *Educational Researcher*, 10(9):9-20.
- Messick, S. 1988. The once and future issues of validity: Assessing the meaning and consequences of measurement. In Wainer, H. & Braun, I.H. (reds.). *Test validity*. Hillside, New Jersey: Lawrence Erlbaum. 33-46.
- Messick, S. 1989. Validity. In Linn, R.L. (red.). *Educational measurement*. New York: MacMillan. 3-103.
- Myburgh, J. 2015. The assessment of academic literacy at pre-university level: A comparison of the utility of academic literacy tests and Grade 10 Home Language results. Ongepubliseerde meestersgraadverhandeling, Universiteit van die Vrystaat. Beskikbaar by <http://hdl.handle.net/11660/2081>.
- Nel, Carryn-Ann. 2014. Lae matriekstandaarde rede vir universiteite wat oorval word met eerstejaars. *Volksblad*, 3 Oktober. 9.



- Noordwes-Universiteit. 2016. *Studeer op die Potchefstroomkampus van die Noordwes-Universiteit*. [Aanlyn]. Beskikbaar by <http://www.nwu.ac.za/af/voornemende-student/studeer-op-potchefstroom>. [8 Desember 2016].
- Oller, J.W. 1979. *Language tests at school: A pragmatic approach*. Londen: Longman.
- O'Sullivan, B. 2011. Language testing. In Simpson, J. (red.). *The Routledge handbook of applied linguistics*. New York: Routledge. 259-273.
- Patterson, R. & Weideman, A. 2013a. The typicality of academic discourse and its relevance for constructs of academic literacy. *Tydskrif vir Taalonderrig*, 47(1):107-123.
- Patterson, R. & Weideman, A. 2013b. The refinement of a construct for tests of academic literacy. *Tydskrif vir Taalonderrig*, 47(1):125-151.
- Pot, A. 2013. Diagnosing academic language ability: An analysis of TALPS. Ongepubliseerde meestersgraadverhandeling, Universiteit van Groningen.
- Pot, A. & Weideman, A. 2015. Diagnosing academic language ability: insights from an analysis of a postgraduate test of academic literacy. *Language Matters* 46(1): 22-43. DOI: <http://dx.doi.org/10.1080/10228195.2014.986665>.

- Raad vir Hoër Onderwys (RHO). 2010. Access and throughput in South African higher education: Three case studies. *Higher Education Monitor* 9. [Aanlyn]. Beskikbaar by [http://www.che.ac.za/media\\_and\\_publications/higher-education-monitor/higher-education-monitor-9-access-and-throughput](http://www.che.ac.za/media_and_publications/higher-education-monitor/higher-education-monitor-9-access-and-throughput). [20 Desember 2014].
- Rademeyer, A. 2005. Minister: Druipelinge kos staat baie. *Die Burger*, 14 Mei. 4.
- Rademeyer, A. 2013. Studente vies oor Engelse vraestelle. *Beeld*, Desember. 13.
- Rajagopalan, K. 2004. The philosophy of applied linguistics. In Davies, A. & Elder, C. (reds.). *The handbook of applied linguistics*. Malden, MA: Blackwell. 397-420.
- Rambiritch, A. 2012. Transparency, accessibility and accountability as regulative conditions for a postgraduate test of academic literacy. Ongepubliseerde doktorsale proefskrif, Universiteit van die Vrystaat. URI: <http://hdl.handle.net/11660/1571>.
- Rambiritch, A. & Weideman, A. 2016. Telling the story of a test: The Test of Academic Literacy for Postgraduate Students (TALPS). In Read, J. (red.). *Post-admission language assessment of university students*. Cham: Springer Publishing International. 197-216. DOI: 10.1007/978-3-319-39192-2.
- Read, J. 2010. Researching language testing and assessment. In Paltridge, B. & Phakiti, A. (reds.). *Continuum companion to research methods in applied linguistics*. Londen: Continuum. 286-300.
- Roets, N. 1990. Afrikaans vir akademiese doeleindes. Ongepubliseerde doktorsale proefskrif. Johannesburg: Randse Afrikaanse Universiteit.

- Schmitt, N. & Celce-Murcia, M. 2010. An overview of applied linguistics. In Schmitt, N. (red.). *An introduction to applied linguistics*, tweede uitgawe. Londen: Hodder Education. 1-15.
- Scott, I., Yeld, N. & Hendry, J. 2007. A case for improving teaching and learning in South African higher education. *Higher Education Monitor 6*. Pretoria: Raad vir Hoër Onderwys. [Aanlyn]. Beskikbaar by [http://www.che.ac.za/sites/default/files/publications/HE\\_Monitor\\_6\\_ITLS\\_Oct2007\\_0.pdf](http://www.che.ac.za/sites/default/files/publications/HE_Monitor_6_ITLS_Oct2007_0.pdf). [20 Desember 2014].
- Sebolai, K. 2016. The incremental validity of three tests of academic literacy in the context of a South African university of technology. Ongepubliseerde doktorsale proefskrif, Universiteit van die Vrystaat. URI: <http://hdl.handle.net/11660/5408>.
- Shohamy, E. 2001. *The power of tests: A critical perspective on the uses of language tests*. Harlow, Essex: Pearson Education.
- Simpson, J. 2011. Introduction. Applied linguistics in the contemporary world. In Simpson, J. (red.). *The Routledge handbook of applied linguistics*. New York: Routledge. 1-7.
- Spack, R. 1997. The acquisition of academic literacy in a second language: A longitudinal study. *Written Communication*, 14(1):3-62.
- Spies, L. 2014. Afrikaans moet weer eie herwin. *Rapport*, 28 Desember. 17.

- Steyn, S. 2013. Compiling an academic word list for Afrikaans. Ongepubliseerde referaat gelewer by die gesamentlike jaarlikse kongres van die Suid-Afrikaanse Vereniging vir Taalonderrig (SAVTO), die Suid-Afrikaanse Vereniging vir Toegepaste Linguistiek (SAALA) en die Linguistevereniging van Suider-Afrika (LVSA). 1-4 Julie, Stellenbosch.
- Steyn, S. 2014. The design and refinement of a test of early academic literacy. Ongepubliseerde meestersgraadverhandeling, Universiteit van Groningen.
- Strydom, N. 2016. Universiteit van Pretoria kies slegs-Engels-taalbeleid. 22 Junie. [Aanlyn]. Beskikbaar by <http://maroelamedia.co.za/nuus/sa-nyuus/universiteit-van-pretoria-kies-slegs-engels-taalbeleid/> [8 Desember 2016].
- Universiteit Stellenbosch, 2016a. *Nuwe taalbeleid: Vrae en antwoorde*. [Aanlyn]. Beskikbaar by <http://www.sun.ac.za/english/Documents/Language/Nuwe%20Taalbeleid%20Vrae%20en%20Antwoorde.pdf>. [18 Junie 2017].
- Universiteit Stellenbosch, 2016b. *Finale taalbeleid*. Junie 2016. [Aanlyn]. Beskikbaar by <http://www.sun.ac.za/english/Documents/Language/Finale%20Junie%20Taalbeleid%20November%202016.pdf>. [18 Junie 2017].
- Van der Slik, F. & Weideman, A. 2005. The refinement of a test of academic literacy. *Per Linguam*, 21(1):23-35.
- Van der Slik, F. & Weideman, A.J. 2008. Measures of improvement in academic literacy. *Southern African Linguistics and Applied Language Studies*, 26(3):363-378.

- Van der Walt, J.L. & Steyn, H.S. 2007. Pragmatic validation of a test of academic literacy at tertiary level. *Ensovoort*, 11(2):138-153.
- Van der Walt, J.L. & Steyn, H.S. 2008. The validation of language tests. *Stellenbosch papers in linguistics*, 38:191-204.
- Van Dyk, T. 2010. Konstitutiewe voorwaardes vir die ontwerp van 'n toets van akademiese geletterdheid. Ongepubliseerde doktorsale proefskrif, Universiteit van die Vrystaat. URI: <http://hdl.handle.net/11660/1918>.
- Van Dyk, T. 2011. Ondersoek na die impak van 'n akademiese geletterdheidsintervensie op eerstejaarstudente se akademiese taalvermoë. *LitNet Akademies*, 8(3). [Aanlyn]. Beskikbaar by <http://www.litnet.co.za/Article/ondersoek-na-die-impak-van-n-akademiese-geletterdheidsintervensie-op-eerstejaarstudente-se>. [20 Desember 2014].
- Van Dyk, T. 2013. Geldigheid vanuit drie paradigmas beskou: 'n Eenheid of veelheid van perspektiewe? *Tydskrif vir Taalonderrig*, 47(1):153-173.
- Van Dyk, T. & Van der Slik, F. 2012. Dalende akademiese geletterdheidsvlakke: Feit of fiksie? *Tydskrif vir Taalonderrig*, 46(2):43-56.
- Van Dyk, T. & Weideman, A. 2004a. Switching constructs: On the selection of an appropriate blueprint for academic literacy assessment. *Tydskrif vir Taalonderrig*, 38(1):1-13.
- Van Dyk, T. & Weideman, A. 2004b. Finding the right measure: From blueprint to specification to item type. *Journal for Language Teaching*, 38(1):15-24.

- Van Rooy, B. & Coetzee-Van Rooy, S. 2015. The language issue and academic performance at a South African university. *Southern African Linguistics and Applied Language Studies*. DOI: 10.2989/16073614.2015.1012691 [14 Desember 2015].
- Van Rooyen, M. 2016. Kovsies wen stryd vir Engels as onderrigtaal. 17 November. [Aanlyn]. Beskikbaar by <http://www.netwerk24.com/Nuus/Onderwys/kovsies-draf-van-2017-af-klas-in-engels-20161117>. [8 Desember 2016].
- Webb, V. 2006. Perspektiewe op moedertaalonderrig. *Tydskrif vir Geesteswetenskappe, Supplement*, 46(2):37-50.
- Wei, L. 2011. From pedagogical practice to critical enquiry. An introduction to applied linguistics. In Wei, L. (red.). *The Routledge applied linguistics reader*. Londen: Routledge. 1-11.
- Weideman, A. 1999. Five generations of applied linguistics: Some framework issues. *Acta Academica*, 31(1):77-98.
- Weideman, A. 2003. Assessing and developing academic literacy. *Per Linguam*, 19 (1&2):55-65.
- Weideman, A. 2006. Transparency and accountability in applied linguistics. *Southern African Linguistics and Applied Language Studies*, 24(1):71-86.
- Weideman, A. 2007a. The redefinition of applied linguistics: Modernist and postmodernist views. *Southern African Linguistics and Applied Language Studies*, 25(4):589-605.

- Weideman, A. 2007b. A responsible agenda for applied linguistics: Confessions of a philosopher. *Per Linguam*, 23(2):29-53.
- Weideman, A. 2009a. Constitutive and regulative conditions for the assessment of academic literacy. *Southern African Linguistics and Applied Language Studies*, 27(3):235-251.
- Weideman, A. 2009b. Uncharted territory: A complex systems approach as an emerging paradigm in applied linguistics. *Per Linguam*, 25(1):61-75.
- Weideman, A. 2011. Academic literacy tests: Design, development, piloting and refinement. *Tydskrif vir Taalonderrig*, 45(2):100-113.
- Weideman, A. 2012. Validation and validity beyond Messick. *Per Linguam*, 28(2):1-14.
- Weideman, A. 2013. Positivism and postpositivism in applied linguistics. In Chapelle, C.A. (red.). *The encyclopaedia of applied linguistics*, volume 7. Oxford: Wiley-Blackwell. 4479-4485.
- Weideman, A. 2014. Persoonlike onderhoud. 17 Desember, Bloemfontein.
- Weideman, A. 2017. *Responsible design in applied linguistics: Theory and practice*. Cham: Springer.
- Weideman, A. (in voorbereiding). A skills-neutral approach to academic literacy assessment. MS. Bydrae tot 'n simposium "Assessing the academic literacy of university students through post-admission assessments", Language Testing Research Colloquium 2017, Bogota, Colombië.

- Weideman, A., Patterson, R. & Pot, A. 2016. Construct refinement in tests of academic literacy. In Read, J. (red.). *Post-admission language assessment of university students*. Cham: Springer Publishing International. 179-196. DOI: 10.1007/978-3-319-39192-2.
- Weideman, A. & Van Rensburg, C. 2002. Language proficiency: Current strategies, future remedies. *Tydskrif vir Taalonderrig*, 36(1 & 2):152-164.
- Xi, X. 2008. Methods of test validation. In Shohamy, E. & Hornberger, N.H. (reds.). *Language testing and assessment. Encyclopaedia of language and education*. New York: Springer. 177-196.
- Yeld, N. et al. 2000. The construct of the academic literacy test (PTEEP). Mimeograph. Kaapstad: Alternative Admissions Research Project, Universiteit van Kaapstad.



## **Abstrak**

Die lae akademiese-geletterdheidsvlakke van studente aan Suid-Afrikaanse universiteite is 'n goeiebekende probleem wat toenemend as een van die hoofoorsake van akademiese onderprestering en swak deurvloeiakoerse bestempel word. Die gebruik van akademiese geletterdheidstoets by tersiêre instansies is 'n direkte gevolg van die behoefte om te bepaal watter studente potensiële risikostudente is wat heel waarskynlik bykomende akademiese ondersteuning gaan benodig ten einde hulle graadkurse binne die voorgeskrewe tydperk met sukses te voltooi.

Een so 'n toets wat reeds 'n geruime tyd lank nasionaal en internasionaal gebruik word om die akademiese geletterdheidsvlak van spesifiek voornemende nagraadse studente te meet, is die Test of Academic Literacy for Postgraduate Students (TALPS). Hierdie toets is egter tans in geen van die amptelike landstale behalwe Engels beskikbaar nie. Gevolglik bevestig hierdie studie die behoefte aan die bevordering van veeltaligheid op nagraadse vlak, selfs in weerwil daarvan dat hierdie strewe toenemend in die huidige universiteitbestel ten gunste van Engels-eentaligheid ondermyn word. Meer spesifiek, beskou hierdie en vorige studies dit as 'n leemte in die gebied van taaltoetsing op tersiêre vlak dat daar tans geen Afrikaanse alternatief vir TALPS beskikbaar is nie.

In antwoord op hierdie behoefte gee hierdie studie uitvoering aan die ontwerp, loodsing en verfyning van 'n nagraadse toets van akademiese geletterdheid in Afrikaans (TAGNaS). Dit word gedoen ooreenkomstig 'n wetenskaplik-teoretiese raamwerk van konstitutiewe en regulatiewe ontwerpbeginsels wat verantwoordelike toetsontwerp in verband bring met, aan die een kant, die inagneming van instrinsieke, konvensionele voorwaardes soos betroubaarheid en geldigheid, en, aan die ander kant, die verrekening van sosiale impakte soos deursigtigheid, verantwoordbaarheid en regverdigheid. Vir die teoretiese begroning van TAGNaS word 'n menigvuldigheid van bewyse ter staving

gebied, waaronder 'n artikulasie van die teoretiese rasionaal waarop die toets gebaseer is ('n kernaspek van konstrugeldigheid), asook 'n beredeneerde opgawe van verskeie statistiese ontledings van die loodstoetsresultate. Aangesien die teoretiese basis van die toets uit 'n onlangs bygewerkte definisie van akademiese geletterdheid bestaan, maak eksperimentering met gewysigde en/of nuwe subtoetse en/of taaktipes deel van hierdie ondersoek uit. Die bevindinge dui daarop dat om vindingryk en kreatief om te gaan met die operasionalisering van 'n konstruk besonder suksesvol en verruimend kan wees, en dui aan dat eksperimentering met die vereistes wat nuwe artikulasies van gevestigde konstrukte stel, geslaag kan wees.

### **Trefwoorde**

Veeltaligheid

Toegepaste linguistiek

Taaltoetsing

Akademiese geletterdheid

Ontwerpbeginsels

Konstitutiewe konsepte

Regulatiewe idees

Betroubaarheid

Geldigheid

Konstrugeldigheid

Konsekwensiële geldigheid

Validasie

## **Abstract**

The low levels of academic literacy of students at South African universities are a well-known problem, and one which is increasingly being identified as one of the prime reasons for a lack of academic performance and low throughput rates. The employment of academic literacy tests by tertiary institutions is a direct result of the need to identify which students are potentially at risk, and would therefore need additional academic support in order to complete their degree courses successfully in the prescribed time.

One such test that has been used for some time nationally and internationally to measure the academic literacy levels specifically of postgraduate students is the Test of Academic Literacy for Postgraduate Students (TALPS). This test, however, is currently unavailable in any of the other official languages. This study therefore confirms the need to promote multilingualism at postgraduate level, even in the face of this goal at present being increasingly undermined by the favouring of a monolingual English dispensation at university. More specifically, this and previous studies acknowledge the lack of an Afrikaans alternative for TALPS as a serious shortcoming in language assessment at tertiary level.

In response to this need the study intends to design, develop, test out and refine a postgraduate of academic literacy in Afrikaans (TAGNaS – Toets van Akademiese Geletterdheid vir Nagraadse Studente). This is done with reference to a theoretical-analytical framework of constitutive and regulative design principles that relate assessment design on the one hand to intrinsic, conventional conditions such as reliability and validity and, on the other, take account of social impacts such as transparency, accountability, and fairness. For the theoretical justification of TAGNaS multiple sets of evidence are presented, among which are the articulation of the theoretical rationale for the test (a key element of construct validity) and an argument integrating several statistical analyses of the

pilot test results. Since the test rests theoretically upon a recently modified definition of academic literacy, experimentation with new or modified subtests and task types was part of the investigation. The results indicate that it can be highly productive to treat the operationalization of the construct with ingenuity and creativity, and also that experimentation with the challenges that new articulations of established constructs provide can be successfully accomplished.

### **Keywords**

Multilingualism

Applied linguistics

Language assessment

Academic literacy

Design principles

Constitutive concepts

Regulative ideas

Reliability

Validity

Construct validity

Consequential validity

Validation

## Bylaag A

### Toetsspesifikasies vir die Test of Academic Literacy for Postgraduate Students (TALPS)

TALPS se bloudruk, soos verskaf deur Du Plessis (2012:53), bevat agt subtoetse wat soos volg met die oorspronklike konstruk van akademiese geletterdheid bely is:

Test section	Aspect of literacy measured
<b>Section 1: Scrambled text</b> A number of sentences that need to be re-organized into a coherent passage.	Textuality (knowledge of cohesion, grammar) Understanding and responding to the communicative function of the text
<b>Section 2: Interpreting graphs and visual information</b> A short text passage and accompanying graph requiring numerical calculations and visual inferences.	Understanding genres Visual literacy Interpreting of information Extrapolation and application of information
<b>Section 3: Academic vocabulary</b> This section includes vocabulary items based on Coxhead's (2000) word list, mainly from the selection of less frequently used words.	Advanced vocabulary knowledge Understanding and responding to the communicative function of the text
<b>Section 4: Text types</b> A selection of phrases and sentences representing different genres which have to be matched with a second group of phrases and sentences.	Understanding genres Identifying registers Making meaning beyond sentence level
<b>Section 5: Understanding texts</b> A lengthy reading passage and series of questions to be answered.	Critical thinking Understanding and responding to the communicative function of the text Deriving meaning beyond sentence level Extrapolating and applying information Distinguishing essential/non-essential information Drawing conclusions and making inferences
<b>Section 6: Grammar and text relations</b> A variation of cloze procedure in which certain words are deleted from a text.	Meaning making Understanding and responding to the communicative function of the text Knowledge of cohesion
<b>Section 7: Text editing</b> A passage in which a number of grammatical errors have been made requiring correction.	Knowledge of syntax Knowledge of morphology Knowledge of semantics
<b>Section 8: Academic writing</b> A short structured essay assignment based on information provided in the test.	Ability to synthesize texts Making meaning beyond the level of the sentence Interpreting information

	Understanding and responding to the communicative function of the text Extrapolation and application of facts Knowledge of genres and registers Applying coherence Referencing
--	--

## Bylaag B

### Toetsspesifikasies vir die Toets van Akademiese Geletterdheid vir Nagraadse Studente (TAGNaS)

Subtoets	Artikulasie met konstruk	Spesifikasies vir toetsontwikkelaars
<p><b>Skommelteks</b></p> <p>Kandidate bepaal die korrekte volgorde van 'n reeks opeenvolgende sinne wat geskommel/deurmekaar aangebied word</p>	<ol style="list-style-type: none"> <li>1. Begrip van samehang; bewustheid van logiese ontwikkeling en teksorganisasie</li> <li>2. Interpretasie van verskillende tekstipes; vatbaarheid vir hulle betekenis</li> <li>3. Skep betekenis verby die sinsvlak</li> <li>4. Begrip van opeenvolging en volgorde</li> <li>5. Begrip van kommunikatiewe funksie</li> </ol>	<p>Rangskikking in die korrekte volgorde</p>
<p><b>Interpretasie van visuele inligting</b></p> <p>Kandidate beantwoord vrae gebaseer op 'n grafiek en doen eenvoudige numeriese bewerkings</p>	<ol style="list-style-type: none"> <li>1. Onderskei tussen hoofsaak en bysaak, feite en menings, stellings en argumente, oorsaak en gevolg</li> <li>2. Klassifiseer, kategoriseer en hanteer data waartussen vergelykings te tref is</li> <li>3. Ekstrapolering en toepassing</li> <li>4. Interpretasie van grafiese of visuele inligting</li> <li>5. Interpretasie van verskillende tekstipes; vatbaarheid vir hulle betekenis</li> <li>6. Skep betekenis verby die sinsvlak</li> </ol>	<p>Tendense</p> <ul style="list-style-type: none"> <li>• Waargenome tendense ten opsigte van volgorde, verhouding en grootte</li> <li>• Voorspellings en skattings gebaseer op tendense</li> <li>• Gemiddeldes oor kategorieë heen, ens.</li> </ul> <p>Verhoudinge</p> <ul style="list-style-type: none"> <li>• Identifiseer en vergelyk verhoudinge uitgedruk i.t.v. breuke of persentasies</li> </ul>

	7. Begrip van opeenvolging en volgorde; doen eenvoudige numeriese skattings en bewerkings	Vergelykings <ul style="list-style-type: none"> <li>• Vergelyk individuele lesings binne 'n kategorie i.t.v. breuke, persentasies of meeteenheid</li> <li>• Vergelyk gekombineerde lesings binne twee/meer kategorieë i.t.v. breuke, persentasies of meeteenheid</li> <li>• Verskille tussen kategorieë</li> <li>• Ekstrapolering en afleidings gebaseer op grafiese of visuele inligting</li> </ul>
<b>Akademie se kennis van akademiese woordeskat</b> Kandidate se kennis van akademiese woordeskat word getoets aan die hand van veelkeusevrae gebaseer op 'n Afrikaanse akademiese woordeskatlys	1. Verstaan en gebruik akademiese woordeskat	Woordeskat in konteks (ook twee woorde weggelaat uit 'n sin i.p.v. slegs een)
<b>Tekstipes</b> Kandidate paar taaluitinge af wat tot dieselfde tekstipe of genre behoort	1. Begrip van samehang 2. Interpretasie van verskillende tekstipes; vatbaarheid vir hulle betekenis 3. Klassifiseer, kategoriseer en hanteer data waartussen vergelykings tref is 4. Skep betekenis verby die sinsvlak	Vergelyk taaluitinge en paar dit af op grond van ooreenstemming t.o.v. tekstipe/genre
<b>Leesbegrip</b> Kandidate beantwoord vrae wat hulle begrip van 'n teks toets	1. Kritiese denke en logiese redenering 2. Onderskei tussen hoofsaak en bysaak, feite en menings, stellings en argumente, oorsaak en gevolg	Primêr <ul style="list-style-type: none"> <li>• Onderskeidings tref</li> <li>• Ekstrapolering, afleiding en toepassing</li> </ul>



<p><b>Vergelyking tussen tekste</b></p> <p>Kandidate beantwoord vrae wat vereis dat hulle die idees in twee tekste met mekaar vergelyk/in verband bring</p>	<ol style="list-style-type: none"> <li>3. Klassifiseer, kategoriseer en hanteer data waartussen vergelykings te tref is</li> <li>4. Tree in wisselwerking met tekste</li> <li>5. Sintetiseer en integreer inligting</li> <li>6. Begrip van samehang; bewustheid van logiese ontwikkeling en teksorganisasie</li> <li>7. Ekstrapolering en toepassing</li> <li>8. Verstaan en gebruik akademiese woordeskat</li> <li>9. Interpretasie van metaforiese taalgebruik; bewustheid van konnotasie, woordspeling en dubbelsinnigheid</li> <li>10. Interpretasie van verskillende tekstipes; vatbaarheid vir hulle betekenis</li> <li>11. Skep betekenis verby die sinsvlak</li> <li>12. Begrip van opeenvolging en volgorde; doen eenvoudige numeriese skattings en bewerkings</li> <li>13. Begrip van kommunikatiewe funksie</li> </ol>	<ul style="list-style-type: none"> <li>• Metaforiese taalgebruik; konnotasie, woordspeling en dubbelsinnigheid</li> <li>• Klassifikasie, kategorisering, vergelyking, sintese en integrasie</li> </ul> <p>Sekondêr</p> <ul style="list-style-type: none"> <li>• Samehang en volgorde</li> <li>• Tekstipes</li> <li>• Numeriese bewerkings</li> </ul>
<p><b>Grammatika en teksverband</b></p> <p>Kandidate moet bepaal waar woorde, asook watter woorde, weggelaat is in 'n teks wat sistematies gemutileer is met behulp van sinsluiting</p>	<ol style="list-style-type: none"> <li>1. Tekstualiteit – samehang en grammatika</li> <li>2. Verstaan en gebruik akademiese woordeskat</li> <li>3. Interpretasie van verskillende tekstipes; vatbaarheid vir hulle betekenis</li> <li>4. Begrip van kommunikatiewe funksie</li> </ol>	<p>'n Wye verskeidenheid aspekte/komponente kan gemeet word; word lukraak bepaal deur die reëlmaat van die mutilasies</p>

<p><b>Opsommingtaak</b></p> <p>Kandidate voltooi 'n grafiese organiseerder wat 'n opsomming van 'n teks bied</p>	<ol style="list-style-type: none"> <li>1. Onderskei tussen hoofsaak en bysaak, feite en menings, stellings en argumente, oorsaak en gevolg</li> <li>2. Begrip van samehang; bewustheid van logiese ontwikkeling en teksorganisasie</li> <li>3. Interpretasie van grafiese of visuele inligting</li> <li>4. Interpretasie van verskillende tekstipes; vatbaarheid vir hulle betekenis</li> <li>5. Skep betekenis verby die sinsvlak</li> <li>6. Tree in wisselwerking met tekste</li> </ol>	<p>Grafiese organiseerder met sleutelwoorde wat hoofidees aandui/uitbeeld, asook verbande tussen hoofidees</p>
<p><b>Verwysings</b></p> <p>Kandidate se kennis van verwysings word getoets aan die hand van vrae gebaseer op 'n bronnelys</p>	<ol style="list-style-type: none"> <li>1. Interpretasie van verskillende tekstipes; vatbaarheid vir hulle betekenis</li> <li>2. Begrip van opeenvolging en volgorde</li> <li>3. Ekstrapolering en toepassing</li> <li>4. Tree in wisselwerking met tekste</li> </ol>	<p>Alfabetiese volgorde</p> <p>Afwyking van konvensie/patroon</p>
<p><b>Akademie skryf</b></p> <p>Kandidate skryf 'n akademiese opstel oor 'n onderwerp wat met die tema van die toets verband hou; tekste wat deel uitmaak van die toets dien as bronmateriaal</p>	<ol style="list-style-type: none"> <li>1. Kritiese denke en logiese redenering</li> <li>2. Onderskei tussen hoofsaak en bysaak, feite en menings, stellings en argumente, oorsaak en gevolg</li> <li>3. Tree in wisselwerking met tekste</li> <li>4. Sintetiseer en integreer inligting</li> <li>5. Begrip van samehang; bewustheid van logiese ontwikkeling en teksorganisasie</li> <li>6. Ekstrapolering en toepassing</li> <li>7. Verstaan en gebruik akademiese woordeskat</li> </ol>	<p>Onderwerp hou tematies verband met inhoud van toets 300 woorde</p> <p>Argumenterend/Beredenerend</p> <p>Verwysings en bronnelys</p> <p>Akademiese skryfkonvensies (formele register en styl, logiese struktuur)</p>

	<ol style="list-style-type: none"><li>8. Interpretasie van verskillende tekstipes; vatbaarheid vir hulle betekenis</li><li>9. Gebruik komplekse grammatikale strukture, leksikale verskeidenheid, formele uitdrukkings en abstrakte/tegniese konsepte</li><li>10. Span lees en skryf vir doeleindes van analise en argumentasie in</li><li>11. Begrip van kommunikatiewe funksie</li><li>12. Skryf op 'n gesaghebbende manier aan 'n denkbeeldige teikenleser</li></ol>	
--	---	--

## Bylaag C: Merkrubriek (TALPS)

CONTENT		Poor	Average	Good	Points
Introduction	Statement of issue – angle to be argued	No clear statement of issue; no point of view to be argued; abrupt or no introduction. <b>(0-1)</b>	States issue and point of view very weakly; not clear what relevance is. <b>(2)</b>	Clearly states issue and point of view, explains relevance and importance. <b>(3)</b>	<b>0-3</b>
Body	Nature of problem/ issue	No or little discussion of the nature of problem/issue, or why it is necessary to deal with it. <b>(0-1)</b>	Unsuccessfully attempts to discuss nature of problem/issue and its importance. <b>(2-3)</b>	Clear discussion of nature of problem/issue, and necessity of addressing it. <b>(4)</b>	<b>0-4</b>
	Discussion of pro's and con's; distinction between fact and opinion, cause and effect; critical thinking	Gives no or little indication that there is more than one side to an argument. No critical argument established.	Attempts to provide both pros and cons, but does so unconvincingly. Unbalanced attention to both fact and opinion and cause and effect.	Provides a comprehensive discussion of possible pro's and con's thereby combining fact and opinion.	
	Argue convincingly for specific point of view	Argumentation is weak, one-sided, unconvincing.	Argument deals with some of the important issues, but not in a convincing way.	Strong, balanced argumentation that leaves the reader convinced of point of view.	
Conclusion	Clearly states again the most important issues	No attempt to highlight again the most important issues in the text. <b>(0-1)</b>	Attempts to again include the most important issues, but does so in an unconvincing and incomplete manner. <b>(2)</b>	Clearly emphasises the main issues again in a structured and non-repetitive manner. <b>(3)</b>	<b>0-3</b>
LANGUAGE AND ORGANISATION		Poor	Average	Good	
Coherence	Introduction	No or little interest in explaining clearly what will follow, or in guiding reader. No transitional words to link to body. <b>(0-1)</b>	Attempts unsuccessfully to frame reader's expectations of what will follow. Weak link to body. <b>(2-3)</b>	Clearly sets out what is to follow, providing a frame for what readers can expect. Clear transition to body. <b>(4)</b>	<b>0-4</b>
	Body	Paragraphs are not linked. No coherence within the argument. Argument does not flow.	Argument patchy in its logic and structure. Paragraphs are weakly linked with transitional words.	Argument is well-structured, paragraphs are linked effectively with transitional words, argument flows.	
	Conclusion	No connection between the issue/thesis introduced in the introduction and what is said in conclusion.	Attempts to restate the issue/thesis, but does so unconvincingly.	Clearly emphasises the thesis again without making it a word by word repetition of the introduction.	
Referencing	No acknowledgement of authorities. No source material is used. No bibliography is present. <b>(0)</b>	Some source material used, but inadequately acknowledged. A bibliography is present. <b>(1)</b>	Appropriate acknowledgement of sources. An accurate bibliography is present. <b>(2)</b>	<b>0-2</b>	
Academic language	No awareness of the communicative functions in academic language. Several lapses in vocabulary. <b>(0)</b>	Some awareness of the various ways of expression in academic language. Contains some lapses in vocabulary. <b>(1)</b>	Aware of the communicative functions and various ways of expression in academic language. Clear use of academic vocabulary. <b>(2)</b>	<b>0-2</b>	
Grammar and spelling	The number of grammatical and spelling errors seriously interferes with the meaning. <b>(0)</b>	Contains some typical errors that could easily have been eliminated. <b>(1)</b>	Primarily error-free and fluent. <b>(2)</b>	<b>0-2</b>	

## Bylaag D: Vertaalde merkrubriek (TAGNaS)

INHOUD		Swak	Gemiddeld	Goed	Punte
Inleiding	Stel die kwessie – standpunt oor saak word ingeneem	Kwessie nie duidelik gestel; geen standpunt ingeneem; abrupte of geen inleiding. <b>(0-1)</b>	Stel kwessie en standpunt baie swak; relevansie onduidelik. <b>(2)</b>	Stel kwessie en standpunt duidelik, verduidelik relevansie en belangrikheid. <b>(3)</b>	<b>0-3</b>
Liggaam	Aard van probleem/kwessie	Geen of beperkte bespreking van aard van probleem/kwessie, of waarom dit aangespreek moet word. <b>(0-1)</b>	Probeer onsuksesvol om aard van probleem/kwessie en belang daarvan te bespreek. <b>(2-3)</b>	Duidelike bespreking van aard van probleem/kwessie, en noodsaaklikheid om dit aan te spreek. <b>(4)</b>	<b>0-4</b>
	Bespreking van argumente en teenargumente; onderskeid tussen feite en menings, oorsaak en gevolg; kritiese denke	Geen of min aanduiding dat 'n argument meer as een kant kan hê. Geen kritiese argumentasie.	Poog om beide argumente en teenargumente te verskaf, maar onoortuigend. Ongebalanseerde aandag aan beide feite en menings en oorsaak en gevolg.	Verskaf uitgebreide bespreking van argumente en teenargumente en kombineer sodoende feite en menings.	
	Betoog oortuigend vir spesifieke standpunt	Argumentasie is swak, eensydig, onoortuigend.	Argument raak enkele belangrike aspekte van kwessie aan, maar onoortuigend.	Sterk, gebalanseerde argumentasie wat die leser van die standpunt oortuig.	
Slot	Duidelike terugkeer na kernaspekte van argument	Geen poging om hoofpunte in die teks uit te lig. <b>(0-1)</b>	Poog om hoofpunte uit te lig, maar op onoortuigende en onvolledige wyse. <b>(2)</b>	Beklemtoon duidelik die hoofpunte op 'n gestruktureerde en nieherhalende manier. <b>(3)</b>	<b>0-3</b>
<b>TAAL EN ORGANISASIE</b>		<b>Swak</b>	<b>Gemiddeld</b>	<b>Goed</b>	
Koherensie	Inleiding	Geen of swak poging om duidelik aan te dui wat gaan volg, of om leser te lei. Geen bindwoorde om aan liggaam te skakel. <b>(0-1)</b>	Probeer onsuksesvol om 'n verwagting by die leser te skep van dit wat gaan volg. Swak geskakel aan liggaam. <b>(2-3)</b>	Sit duidelik uiteen wat gaan volg, verskaf 'n raamwerk van dit wat lesers kan verwag. Duidelike oorgang na liggaam. <b>(4)</b>	<b>0-4</b>
	Liggaam	Paragrafe is nie geskakel. Geen samehangende argument. Argument vloei nie.	Argument toon gebreke t.o.v. logika en struktuur. Paragrafe is swak geskakel. Swak gebruik van bindwoorde.	Argument is goed gestruktureer, paragrafe is doeltreffend geskakel d.m.v. bindwoorde, argument vloei.	
	Slot	Geen skakeling tussen die kwessie/tesis in die inleiding en die inhoud van die slot.	Probeer om die kwessie/tesis in ander woorde te herbesoek, maar doen dit onoortuigend.	Beklemtoon die tesis weer duidelik sonder om dit woord-vir-woord uit die inleiding te herhaal.	
Verwysing	Geen erkenning aan bronne. Geen bronmateriaal is gebruik. Geen bronnelys is teenwoordig. <b>(0)</b>	Van die bronmateriaal is gebruik, maar onvoldoende erken. 'n Bronnelys is teenwoordig. <b>(1)</b>	Gepaste en voldoende erkenning aan bronne. 'n Korrekte bronnelys is teenwoordig. <b>(2)</b>	<b>0-2</b>	
Akademiese taal	Toon geen bewustheid van die kommunikatiewe funksies in akademiese taal. Verskeie swak woordeskatkeuses. <b>(0)</b>	Toon redelike bewustheid van die uitdrukkingswyses in akademiese taal. Enkele swak woordeskatkeuses. <b>(1)</b>	Toon bewustheid van kommunikatiewe funksies en uitdrukkingswyses in akademiese taal. Duidelike gebruik van akademiese woordeskat. <b>(2)</b>	<b>0-2</b>	
Grammatika en spelling	Die hoeveelheid grammatikale en spelfoute lei tot ernstige onduidelikhede t.o.v. betekenis. <b>(0)</b>	Bevat enkele tipiese foute wat maklik uitgeskakel kon word. <b>(1)</b>	Hoofsaaklik foutvry en vlot. <b>(2)</b>	<b>0-2</b>	

## Bylaag E

### Potensiële probleemitems in die verfynde weergawe van TAGNaS

#### Eerste toetsfase

Item ID	P / Item Mean	R	Flag(s)
6	0.934	0.260	HP
11	0.517	0.087	K, LR
26	0.686	0.100	LR
31	0.863	0.370	HP

#### Tweede toetsfase

Item ID	P / Item Mean	R	Flag(s)
9	0.535	0.095	K, LR
10	0.152	0.049	LR
13	0.172	0.042	K, LR
14	0.545	0.112	LR
15	0.212	-0.191	K, LR
25	0.515	0.138	LR
40	0.525	0.023	K, LR
41	0.273	0.148	LR
48	0.232	0.100	K, LR
53	0.354	0.136	LR
74	0.869	0.366	HP
76	0.889	0.339	HP
78	0.869	0.259	HP
81	0.859	0.389	HP
83	0.859	0.302	HP
87	0.545	0.057	K, LR